PLOS ONE

Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces.

--Manuscript Draft--

Manuscript Number:	PONE-D-15-09499R1
Article Type:	Research Article
Full Title:	Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces.
Short Title:	Categorical Perception of Emotions in Face Parts.
Corresponding Author:	Martin Wegrzyn Universitat Bielefeld Bielefeld, GERMANY
Keywords:	Faces; emotion; expression; categorical perception
Abstract:	Human observers are remarkably proficient at recognizing expressions of emotions and at readily grouping them into distinct categories. When morphing one facial expression into another, the linear changes in low-level features are insufficient to describe the changes in perception, which instead follow an s-shaped function. Important questions are, whether there are single diagnostic regions in the face that drive categorical perception for certain parings of emotion expressions, and how information in those regions interacts when presented together. We report results from two experiments with morphed fear-anger expressions, where (a) half of the face was masked or (b) composite faces made up of different expressions were presented. When isolated upper and lower halves of faces were shown, the eyes were found to be almost as diagnostic as the whole face, with preserved categorical perception. In contrast, the mouth allowed for a substantially lesser amount of accuracy and nearly absent categorical perception. When a composite face consisting of mismatched upper and lower halves was used and observers were instructed to exclusively judge either the expression of mouth or eyes, the to-be-ignored part always influenced perception of the target region. In line with experiment 1, the eye region exerted a much stronger influence on mouth judgements than vice versa. Again, categorical perception was significantly more pronounced for upper halves of faces. The present study shows that identification of fear and anger in morphed faces relies heavily on information from the upper half of the face, particularly the eye region. Categorical perception is possible when only the upper face half is present, but compromised when only the lower part is shown. Moreover, observers tend to integrate all available features of a face, even when trying to focus on only one part.
Order of Authors:	Martin Wegrzyn
	Isabelle Bruckhaus
	Johanna Kissler
Opposed Reviewers:	
Response to Reviewers:	Reviewer #1: As the author said in Discussion, their result was obtained with angry and fearful images, it may not be able to generalize to other expressions, such as happy, surprised or sad. This limitation should be reflected in the title and abstract. Thank you for your suggestion. We have changed the title of our manuscript to „Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces.“. Also, there are two passages in the abstract which make clear that our study deals with fearful and angry faces only: „We report results from two experiments with morphed fear-anger expressions [...]“

(abstract, 3rd paragraph)
„The present study shows that identification of fear and anger in morphed faces relies heavily on information from the upper half [...]"(abstract, last paragraph)
In addition, we have added the following clarification to the abstract:
„Important questions are, whether there are single diagnostic regions in the face that drive categorical perception for certain parings of emotion expressions [...]" (abstract, 2nd paragraph)

The main argument of this paper, as stated in Section 2.4, is based on the assumption: "The linear function (Flinear(x;α,β)=β*x+α) should correspond best to changes in low-level features and would be a good fit if observers' performance closely followed the basic image properties. The logistic function (Flogistic(x;α,β)=1/[1+exp(-β(x-α))]) is well suited to describe observers' performance in psychometric terms [25, p. 82] and should provide an optimal fit if the observers engage in categorical perception". The authors made this assumption without an explanation or a reference (the cited Kingdom & Prins book was about logistic function, not its implications). Actually, this assumption is wrong. The logistic psychometrics function occurs for any linear system limited by Gaussian noise (see Green & Swets, 1966, Signal Detection Theory; Duda & Hart, 1973, Pattern Classification and Scene Analysis, and the Kingdom & Prins book cited by the authors), even the one whose decision is based on simple image features, such as contrast (e.g., Foley & Legge, 1981, Vision Res., 21, 1041-53). Hence, whether the psychometric function is linear or logistics cannot help the authors to decide whether the observer is based on "low-level features" or "categorical perception" (By the way, neither term were defined in the paper).

We agree and would like to thank the reviewer for pointing out this important issue. There certainly is no one-to-one mapping between categorical perception and an s-shaped response function. However, we believe that the presence of an s-shaped logistic function is a necessary condition to claim categorical perception. For example, the assumption of categorical perception for syllables (e.g. ba/pa) or colours would be incompatible with a strictly linear increase in the response function. Although the logistic function can arise due to different reasons, as pointed out by the reviewer, its absence would provide strong evidence against categorical perception (i.e. in the sense of the modus tollens).
Accordingly, we wrote in our manuscript:
"[...] the linear changes in low-level features are insufficient to describe the changes in perception, which instead follow an s-shaped function"
and
"The linear function […] would be a good fit if observers' performance closely followed the basic image properties […] The logistic function [...] should provide an optimal fit if the observers engage in categorical perception"
We believe that these statements are logically sound, as they do not claim that we want to dichotomize between categorical perception and a strict following of low-level features. There was a statement in the results section which we agree was problematic:
"Results show that the slopes were steepest for the whole face condition […] Thus, categorical perception was strongest when all face features were visible [...]"
This too far-reaching claim has been removed from the manuscript.
Also, we have added a discussion of these issues in the methods section of our manuscript:
"It should be noted that the logistic function might occur for any linear system limited by gaussian noise [26], including decisions based on simple image features such as contrast [27]. However, while the mere presence of a nonlinear or s-shaped curve in an identification task is not sufficient to claim the occurrence of categorical perception, its absence argues strongly against it." (2.4 Analysis, 2nd paragraph)

We have also added explicit definitions of the term "low-level features" to the introduction of our manuscript (in the respective paragraph, a definition of "categorical perception" was already given):
"Such categorical perception is demonstrated by identification tasks where linear changes in low-level features (e.g. basic visual properties like colour or shape) lead to non-linear changes in perception, which are best described by a sigmoid function with a steep category boundary. In general, categorical perception refers to the phenomenon that objects from the same category will be perceived as more similar

than objects from different categories [5], despite being equally far apart from each other on a given physical dimension." (Introduction, 3rd paragraph)

The second reason why a comparison between linear and logistic fits is meaningless is that any linear psychometric function can be considered as an approximation to a logistic function with limited range.
We agree that our initial analyses with direct comparisons between the fit of a logistic and linear function should be omitted. As it still is of importance to us to provide objective metrics for how s-shaped the logistic function is for each condition, we now adopt the rationale of comparing the degree of superiority a logistic fit between conditions. This avoids attaching meaning to the fact that the logistic function per se provides a superior fit over the linear one. The metric of interest instead is, for which condition this superiority is most pronounced. We feel that this is an important reframing of our analysis rationale and hope that the reviewer agrees.
This amended analysis rationale is made explicit in the methods section:
"[…] while the mere presence of a nonlinear or s-shaped curve in an identification task is not sufficient to claim the occurrence of categorical perception, its absence argues strongly against it. Therefore, a linear function ($F_{linear}(x;\alpha,\beta)=\beta*x+\alpha$), as an approximation to a logistic function with uniform slope was fitted to the data. If the logistic and linear function provide a similar fit for a certain condition, this would indicate that observers are unable or at least very poor at distinguishing between two distinct categories." (2.4 Analysis, 2nd paragraph)
The results section has been amended, with the respective sections only reporting the degree of superiority of the logistic function between conditions. Figures 3 and 6 have also been modified accordingly (cf. sections 3.2.3 -3.2.5 for experiment 1 and sections 3.2.3 -3.2.5 for experiment 2) .

The authors also made a considerable effort describing the slope of the psychometric function. Their rationale was described in Section 4.2 "...slope information (β parameter), to quantify how steep the category boundary is for a certain condition." There are many reasons underlying a specific slope of a psychometric function. A shallow slope may be caused by a large variance of the underlying noise distribution (Ashby & Townsend, 1986, Psych. Rev., 93, 154-79) or a low level of uncertainty in the system (Pelli, 1985, JOSAA, 2, 1508-32; Tyler & Chen, 2000, Vision Res., 40, 3121-44). It is not necessary reflecting the steepness of the boundary.

We agree that alternative explanations should always be considered. Therefore, we added the reviewer's concerns to the discussion of our manuscript:
"The employed identification task could be complemented by a discrimination task. There, it should be easier to discriminate two images when they are on opposing sides of the category boundary, compared with being both sampled from within the same category [5,6]. Thus, the sigmoid curves from the present experiments could be used to predict the discrimination performance in a follow-up experiment. As differences in slope between conditions can reflect differences in variance of the underlying noise distribution or the uncertainty of the observer [34,35], it would be important to establish that a between-category advantage [5] is indeed present at the point where the psychometric function is steepest in the identification task." (Discussion section, 9th paragraph)

However, there are two main reasons why we are confident that our use of slope parameters is sensible:
Firstly, there is a compelling relationship between the steepness of the slope in Exp. 1 and the performance in the composite face task in Exp. 2. Namely, the steeper the slope in Exp. 1, the more the respective face part is resilient to biasing influence. Likewise, a shallow slope in Exp. 1 leads to a strong susceptibility to distractors in Exp. 2. Therefore, the steepness of the slope indicates how robustly observers can distinguish between two categories in the presence of distractors.
Secondly, there is a substantive body of literature linking the steepness of the slope to performance in a discrimination task, e.g. an AB-X or same-different task (Etcoff & Magee 1992 Cognition; Calder et al. 1996 Visual Cognition, Young et al. 1997 Cognition). In the discussion section of our manuscript, we addressed the need to validate the current data with a discrimination task:
"On the basis of the present data from experiment 1, we would therefore predict that discrimination performance at the category boundary should be higher for eye than

mouth judgements." (Discussion section, 9th paragraph)
Since our submission of the current manuscript, we have acquired pilot data from a three-alternatives forced-choice discrimination task. In this task, three faces are shown with one oddball face differing in its expression by 10%. While we are still in the process of designing the task and the available data are only tentative, they fit our predictions. They illustrate that with full face and upper face halves observers show best performance for the middle morphs, where the psychometric curve is steepest, while discrimination performance is uniformly low for the lower face half (cf. https://www.dropbox.com/s/45nxm9uzaxwjyow/FigureR1.png?dl=0), where the slope is shallow. The discrimination performance seems to closely mimic the steepness of the curves in the current identification task.

We feel that these results help to support our claim that the steepness of the slope in an identification task carries valuable information about the ability of an observer to engage in categorical perception.

Statistical procedures for model comparison and cross validation were not described. While the authors provided Python codes in the Supplementary material, it is not enough. Not everyone uses Python. The authors should describe their methods in the text such that the readers can get an idea about the method they used without consulting supplementary material.

We apologize if the methods section was not comprehensive enough. It was not our intention that readers would have to study the supplement to properly understand the statistical procedures. We have now added the following paragraphs to the methods section:
Firstly, to better explain the use of sums of squared errors for model comparison:
"Functions were fitted using an iterative optimization strategy, employing a non-linear least-squares procedure to derive the parameters for intercept and slope which describe the best-fitting function for the raw data. These parameters, together with the sum of squared errors (ss^2) of the best-fitting linear or logistic function were then compared between conditions." (2.4 Analysis, 3rd paragraph)

Secondly, to explain the cross-validation the following paragraph was extended:
"Furthermore, for cross-validation, data from one half of the experiment were used to predict the other half of the data. Since each participant performed every trial twice, data from the first half of the experiment were used to predict the data from the second half (and vice versa). For each condition, one half of the participants' data was used to fit a logistic or linear function. Subsequently, the ss^2 between these data and the left-out half were computed. For example, a logistic function was fit to the first half of the whole face condition and this fitted function was then compared to the raw data of the second half of the experiment, using ss^2." (2.4 Analysis, 4th paragraph)

Reviewer #2: This manuscript addresses a topical research question and provides interesting support for existing knowledge about the importance of both analytic and holistic processing of faces for facial expression categorization.

General comment: As acknowledged by the authors, evidence for categorical perception of facial expressions is numerous. The authors should highlight more explicitly what their findings add to existing knowledge and theories. In the introduction, theory-based predictions could be made.

We have made our goals and predictions more explicit in the introduction:
"Previous masking studies with full-blown fear and anger expressions have shown that both expressions are mainly recognized from the eye region [13] and studies with neurological patients also point to a special role of the eyes for recognizing fear [22]. Therefore, one can hypothesize that categorical perception of fear and anger will rely mostly on the features in the upper half of the face. In line with research on the composite face illusion [19,20], the more informative half of the face should also dominate the perception of a full face." (Introduction, paragraph before last)

We have also completely re-worked the last paragraphs of the discussion section to describe the value of the findings for our theoretical understanding of face perception:
"Expanding previous work, the study shows that observers can perform categorical

perception even when viewing only a single face part. For the employed fear-anger pairings, the eye region was sufficient to allow for decisions with a high degree of certainty. This demonstrates that categorical perception can emerge based on single features only, adding to our understanding of how categorization works in complex stimuli such as faces.
Even though the eye region was almost as diagnostic as the whole face, it was not immune to biasing influences in a composite face task. When a full face with conflicting expressions was shown, the psychometric curves for eye judgements were shifted in accordance with the expression in the lower half of the face. For the first time, we can therefore show that holistic perception of a full face can be explained based on the diagnostic value of its constituent parts. In addition, given that observers were aware that a holistic processing strategy impaired their performance, this illustrates that humans involuntarily process faces as a gestalt when trying to make categorical decisions about its expression." (Discussion Section, last paragraph)

Statistical analysis: The use of a large number of t-tests without correction for inflated Type 2 error is not appropriate. The data in most sections can be analysed using ANOVA (excluding the curve-fit analyses at the beginning of sections 2.5.3 and 3.2.3) and post-hoc corrections should be used for analysis of significant main or interaction effects. An indication of effect size of effects should also be included (eta squared).

Thank you for this suggestion. We now provide ANOVA analyses and hope that readers will benefit from them. However, we would like to point out that all information provided by an ANOVA was already conveyed in the original analyses. Each section of the results now features analyses of a repeated-measures ANOVA as well as the original pairwise tests for more focused and directional testing.
Changes have been made throughout the results section and are marked in red color. One example of the use of repeated-measures ANOVA is given here:
"Identification performance was compared across all 11 morphing grades for the three masking conditions (full face, upper face half, lower face half). Participants' performance is illustrated in Fig. 2. A 3x11 repeated-measures ANOVA showed a significant main effect of morphing grade, reflecting that anger responses increase as the faces are morphed towards this expression ($F(10,270)=620.7$, $p<0.001$, $\eta p^2=0.96$). There was no main effect of masking condition ($F(2,54)=2.3$, $p=0.112$, $\eta p^2=0.08$), indicating that there is no bias that would shift the responses between conditions. A significant condition-by-morphing grade interaction ($F(20,540)=63.5$, $p<0.001$, $\eta p^2=0.70$) indicated that the shape of the response curve differs between conditions." (Section 2.5.1, first paragraph)

We have performed a large number of paired t-tests, due to the fact that we have a very fine-grained parametric modulation of our independent variable (11 morphing steps). We believe that it is of interest to provide the readers with precise information for which morphing grades the masking conditions differ. Therefore, we think that the t-test do add value to the manuscript and help the reader in interpreting the figures.
There are two kinds of corrections we could do. A correction for Type 1 error (e.g. a Bonferroni-correction), would basically penalise any design with fine-grained modulation of an independent variable. As it is not the case that the statistical evidence for an effect weakens as more and more significant differences are found, we find this inappropriate. If two conditions differ in 11 out of 11 levels of a variable (in a systematic and plausible pattern), this alone must be considered a highly significant effect under a binomial distribution (e.g. like a coin coming up heads 11 times in a row).
A correction for Type 2 error would mean that we should increase our alpha to be more sensitive to smaller differences (e.g. by using a p-value higher than 0.05). We believe that the best we can do to avoid Type 2 error is to have a good signal-to-noise ratio and collect a large sample. We feel that this is the case in both of our experiments, given that (a) the effects are robust and visible on a single subject level and (b) the effects can be replicated when splitting the data in half.
Therefore, while we share the concerns about both Type-1 and Type-2 errors with the reviewer, we are convinced that our design, our comprehensive analyses and the results speak in favour of robust and reproducible effects.

There are a few sentences that seem out of context or seem non-informative (e.g. Discussion, second page, par 4: "Anonymization of pictures…").

	We would like to thank the reviewer for pointing this out. We have re-phrased the referenced part to make our point clearer and establish a better connection to the previous paragraph: "Regarding identity recognition, there are many studies that demonstrate the importance of the eyes. For example, when a face has to be learned in a training run and later distinguished from distractors differing in one feature, it is easier to recognize identity changes due to differences in the eyes, compared to nose or mouth changes [29]. Masking out the eyes may even represent a legal requirement in some contexts for ensuring anonymity. Although the scientific basis for the effectiveness of this procedure is equivocal at best [cf. 30], it illustrates the paramount importance we intuitively place on the eye region. " (Discussion Section, 6th paragraph) Discussion: Similar to the first point: The findings could have been compared more critically to findings of previous studies and the theoretical contribution could be highlighted more explicitly. The recommendations for future studies (e.g. which features are relevant for different emotions) were interesting, but it is not argued clearly what this would add to theoretical understanding. We are glad that the reviewer found our discussion to cover interesting material. We have further added more elaborate explanations regarding the theoretical significance of our suggestions. We have added two major points to our discussion, also inspired by the study by Pollux et al. 2014 which the reviewer kindly drew our attention to: "[...]regarding the results of the present experiments, we would predict that the basic mechanisms outlined here will be replicated with other facial expressions, but perhaps not for the same features. For example, happy-surprised pairings should be best recognized from the mouth, as both involve mainly muscles of the lower face half [10,13,32]. Interestingly, eye-tracking studies found that more fixations on the eye regions correlates with improved emotion recognition for all expressions [14,33] and argue that fixating the eyes reflects a holistic processing strategy. One could test this assumption using happy-surprise pairings and the composite face task from experiment 2. If holistic processing is associated with fixating the eyes, then eye judgements should be easier in a composite face task, even if a strongly diagnostic distracting lower face half is present." (Discussion Section, 8th paragraph) And we have further clarified why the use of a discrimination task is warranted in future studies investigating the phenomenon: "The employed identification task could be complemented by a discrimination task. There, it should be easier to discriminate two images when they are on opposing sides of the category boundary, compared with being both sampled from within the same category [5,6]. Thus, the sigmoid curves from the present experiments could be used to predict the discrimination performance in a follow-up experiment. As differences in slope between conditions can reflect differences in variance of the underlying noise distribution or the uncertainty of the observer [34,35], it would be important to establish that a between-category advantage [5] is indeed present at the point where the psychometric function is steepest in the identification task." (Discussion Section, 9th paragraph) Discussion of training facial expression categorization and the role of gaze focus could have been linked to a few existing studies examining related topics (e.g. Dadds, 2012; Pollux et al, 2014). More thought could be given to structure of the whole discussion section. Thank you for the helpful references. In addition to our discussion of the Pollux et al. 2014 paper which we reference in the comment above, we now also discuss this study and its high relevance in our discussion of face recognition training: "[...] Another interesting question is to what degree the failure to concentrate on one half of a composite face can be voluntarily overridden, especially since previous work has shown that emotion recognition improvements are associated with increased fixation of the eyes [33]. This could be investigated by introducing training blocks before the experiment proper, a forced minimum exposure time to discourage spontaneous responses, biofeedback from an eye-tracking device helping to fixate on a specific face part, and rewards for correct responses." (Discussion, 2nd paragraph before last)

Additional Information:	
Question	**Response**
Financial Disclosure Please describe all sources of funding that have supported your work. A complete funding statement should do the following: Include **grant numbers and the URLs** of any funder's website. Use the full name, not acronyms, of funding institutions, and use initials to identify authors who received the funding. **Describe the role** of any sponsors or funders in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. If they had no role in any of the above, include this sentence at the end of your statement: "*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*" If the study was **unfunded**, provide a statement that clearly indicates this, for example: "*The author(s) received no specific funding for this work.*" * typeset	Research was funded by the Deutsche Forschungsgemeinschaft (DFG; www.dfg.de), Cluster of Excellence 277 "Cognitive Interaction Technology", in form of a scholarship awarded to MW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Competing Interests You are responsible for recognizing and disclosing on behalf of all authors any competing interest that could be perceived to bias their work, acknowledging all financial support and any other relevant financial or non-financial competing interests. Do any authors of this manuscript have competing interests (as described in the PLOS Policy on Declaration and Evaluation of Competing Interests)? **If yes**, please provide details about any and all competing interests in the box below. Your response should begin with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests:*	The authors have declared that no competing interests exist.

If no authors have any competing interests to declare, please enter this statement in the box: "*The authors have declared that no competing interests exist.*" * typeset	
Ethics Statement You must provide an ethics statement if your study involved human participants, specimens or tissue samples, or vertebrate animals, embryos or tissues. All information entered here should **also be included in the Methods section** of your manuscript. Please write "N/A" if your study does not require an ethics statement. **Human Subject Research (involved human participants and/or tissue)** All research involving human participants must have been approved by the authors' Institutional Review Board (IRB) or an equivalent committee, and all clinical investigation must have been conducted according to the principles expressed in the Declaration of Helsinki. Informed consent, written or oral, should also have been obtained from the participants. If no consent was given, the reason must be explained (e.g. the data were analyzed anonymously) and reported. The form of consent (written/oral), or reason for lack of consent, should be indicated in the Methods section of your manuscript. Please enter the name of the IRB or Ethics Committee that approved this study in the space below. Include the approval number and/or a statement indicating approval of this research. **Animal Research (involved vertebrate animals, embryos or tissues)** All animal work must have been conducted according to relevant national and international guidelines. If your study involved non-human primates, you must provide details regarding animal welfare	The study was approved by the ethics board of Bielefeld University (Ethic Statement Nr. 2014-010). All participants gave oral informed consent before taking part in the experiments.

and steps taken to ameliorate suffering; this is in accordance with the recommendations of the Weatherall report, "The use of non-human primates in research." The relevant guidelines followed and the committee that approved the study should be identified in the ethics statement. If anesthesia, euthanasia or any kind of animal sacrifice is part of the study, please include briefly in your statement which substances and/or methods were applied. Please enter the name of your Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board, and indicate whether they approved this research or granted a formal waiver of ethical approval. Also include an approval number if one was obtained. **Field Permit** Please indicate the name of the institution or the relevant body that granted permission.	
Data Availability PLOS journals require authors to make all data underlying the findings described in their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found. Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. **Please note that simply stating 'data available on request from the author' is not acceptable.** ***If, however, your data are only available upon request from the author(s), you must answer "No" to the first question below, and explain your exceptional situation in the text box provided.*** Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	Yes - all data are fully available without restriction

Please describe where your data may be found, writing in full sentences. **Your answers should be entered into the box below and will be published in the form you provide them, if your manuscript is accepted.** If you are copying our sample text below, please ensure you replace any instances of XXX with the appropriate details. If your data are all contained within the paper and/or Supporting Information files, please state this in your answer below. For example, "All relevant data are within the paper and its Supporting Information files." If your data are held or will be held in a public repository, include URLs, accession numbers or DOIs. For example, "All XXX files are available from the XXX database (accession number(s) XXX, XXX)." If this information will only be available after acceptance, please indicate this by ticking the box below. If neither of these applies but you are able to provide details of access elsewhere, with or without limitations, please do so in the box below. For example: "Data are available from the XXX Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data." "Data are from the XXX study whose authors may be contacted at XXX." * typeset	All relevant data are within the paper and its Supporting Information files.
Additional data availability information:	

Universität Bielefeld **Department of Psychology**

University of Bielefeld | P.O. Box 10 01 31 | 33501 Bielefeld

Kun Guo
University of Lincoln
Brayford Pool, Lincoln
Lincolnshire, LN6 7TS
United Kingdom

Martin Wegrzyn

Room: T5-213
Tel.: ++49 521-106-4527
Fax: ++49 521.106-6422
martin.wegrzyn@uni-bielefeld.de
www.uni-bielefeld.de/psychologie/ae/AE02/

Bielefeld, 15 May 2015

Resubmission of Manuscript (PONE-D-15-09499)

Dear Editor,

enclosed please find the revised version of our manuscript entitled "Categorical Perception of Emotion Expressions in Whole, Masked and Composite Faces".

We would like to thank you and the reviewers for the helpful comments on our study. We have adressed all issues raised by the reviewers and have attached a point-by-point answer to their comments. The manuscript itself has been thorougly revised, and all changes are marked in red color.

We feel that the suggestions have helped to greatly improve the manscript and we hope that you and the reviewers feel that same way. We would appreciate a positive review and are looking forward to your reply.

Sincerely for the authors,

M. Wegrzyn

Martin Wegrzyn

Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces.

Martin Wegrzyn[1,2,*], Isabelle Bruckhaus[1] & Johanna Kissler[1,2]

[1]Department of Psychology, Bielefeld University, Bielefeld, Germany

[2]Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Bielefeld, Germany

Manuscript Requirements:

Title: 90 Characters

Abstract: 283 Words

Text: 5498 Words

References: 39

Tables: 0

Figures: 7

***Corresponding Author:**

Martin Wegrzyn

Bielefeld University

Department of Psychology

Postfach 10 01 31

33501 Bielefeld

Tel. +49 521-106-4527

Fax. +49 521-106-154454

martin.wegrzyn@uni-bielefeld.de

ABSTRACT

Human observers are remarkably proficient at recognizing expressions of emotions and at readily grouping them into distinct categories. When morphing one facial expression into another, the linear changes in low-level features are insufficient to describe the changes in perception, which instead follow an s-shaped function.

Important questions are, whether there are single diagnostic regions in the face that drive categorical perception for certain parings of emotion expressions, and how information in those regions interacts when presented together.

We report results from two experiments with morphed fear-anger expressions, where (a) half of the face was masked or (b) composite faces made up of different expressions were presented.

When isolated upper and lower halves of faces were shown, the eyes were found to be almost as diagnostic as the whole face, with preserved categorical perception. In contrast, the mouth allowed for a substantially lesser amount of accuracy and nearly absent categorical perception.

When a composite face consisting of mismatched upper and lower halves was used and observers were instructed to exclusively judge either the expression of mouth or eyes, the to-be-ignored part always influenced perception of the target region. In line with experiment 1, the eye region exerted a much stronger influence on mouth judgements than vice versa. Again, categorical perception was significantly more pronounced for upper halves of faces.

The present study shows that identification of fear and anger in morphed faces relies heavily on information from the upper half of the face, particularly the eye region. Categorical perception is possible when only the upper face half is present, but

compromised when only the lower part is shown. Moreover, observers tend to integrate all available features of a face, even when trying to focus on only one part.

1. INTRODUCTION

Facial expressions are a powerful means of conveying information about the emotional state of an individual. Recognizing and correctly interpreting these non-verbal signs is of vital importance for successful social interaction [1].

Faces are not only complex stimuli, but their expressions can often be subtle and ambiguous. Despite this, human observers can readily recognize expressions of emotions and assign them distinct labels. A large body of research suggests that there are a number of basic emotions which are correctly recognized by most observers [2], although their exact number is under debate [3]. Even when the expressions' intensity varies continuously, as in manipulated faces where one expression is gradually morphed into another, human observers readily group them into distinct categories with a high level of confidence [4].

Such categorical perception is demonstrated by identification tasks where linear changes in low-level features (e.g. basic visual properties like colour or shape) lead to non-linear changes in perception, which are best described by a sigmoid function with a steep category boundary. In general, categorical perception refers to the phenomenon that objects from the same category will be perceived as more similar than objects from different categories [5], despite being equally far apart from each other on a given physical dimension. One example is human colour perception, where two different wavelengths will only be perceived as two different colours when a category boundary is crossed [6]. For facial expressions, categorical perception has been demonstrated in a number of seminal studies [7–9].

Faces are inherently multidimensional stimuli, and in face processing categorical perception can relate to different dimensions (identity, age, attractiveness, gender, emotion), which in turn may depend on different facial features. Regarding emotion, each expression of a basic emotion can be described in terms of a number of muscle groups that are active when a person shows that particular emotion [10]. For example, fear will most frequently be expressed in the face by the brows raised and drawn together, the upper eyelids raised, the lower eyelids tensed, the lips stretched back and, in some cases, the mouth opened [11,12]. Furthermore, masking studies have demonstrated that each basic emotion is recognized through different diagnostic areas of the face [13]. For example, fear can be best recognized when the eyes are visible, while happiness is best inferred from the mouth region. On the other hand, eye tracking studies have shown that for all six basic expressions, including ones of low intensity, people focus primarily on the eyes [14]. While eye tracking studies shed light on how observers usually inspect faces under free viewing conditions, masking studies help to understand what information is indispensable in order to make a correct decision. Therefore a masking approach helps to make causal inferences about the diagnostic value of each facial feature.

To understand categorical perception of facial expressions, it is essential to know which information in the face is used to make a categorical decision and how facial features are integrated into a perceptive whole, so that successful categorization can occur.

The term 'holistic processing' refers to the fact that faces are preferentially processed as a whole [15–17]. Findings from the composite face illusion, where complementary upper and lower face halves are combined into a whole face, serve to illustrate this mechanism: It is considerably more difficult to make a decision about the properties of either half in such composites than it is to judge each part in isolation [18]. It may be more difficult to recognize a person's identity from a lower face half when it has been combined with the upper face half of another person, while the same task is easy when each half is presented in isolation [18]. Using happy and angry faces, interference from composites with mis-matching expressions have been found reflected in slower reaction times [19,20].

So far, both masking studies [13] and composite face paradigms [19] used full-blown facial expressions instead of gradually morphed faces. Chen and Chen [21] have used morphed happy-sad expressions in a composite face, but always asked to judge the whole face and not its parts, therefore making it difficult to draw inferences on how the perception of one facial feature is influenced by another one.

Masking or composite face studies with morphed faces however allow to better describe human behaviour in psychophysical terms, as the subjective perception dissociates from the changes in low-level features. They also allow to better understand how we process subtle and ambiguous expressions, which are arguably most relevant in daily life. Therefore, it is useful to investigate the relationship of categorical perception and featural (i.e. focusing on single parts) or configural (i.e. integrating parts into a whole) face processing with masked and composite faces. Main questions are, whether there are diagnostic regions in the face that are sufficient to allow for categorical perception, and whether multiple features need to be integrated when making categorical judgements. Therefore, to better understand mechanisms of categorical perception of facial expressions, a promising approach is to break down the face into its features and understand how each contributes to identifying a certain expression. In a second step, the features can be re-assembled into a whole and one can investigate how they are integrated to infer an expression from a whole face.

To address these questions, two experiments with faces morphed from fear to anger were performed. Beginning with the seminal work by Etcoff and Magee [9], fear-anger pairings have been frequently used for investigating categorical perception (e.g. [7,8] ; cf. [5]). Previous masking studies with full-blown fear and anger expressions have shown that both expressions are mainly recognized from the eye region [13] and studies with neurological patients also point to a special role of the eyes for recognizing fear [22]. Therefore, one can hypothesize that categorical perception of fear and anger will rely mostly on the features in the upper half of the face. In line with research on the composite face illusion [19,20], the more informative half of the face should also dominate the perception of a full face.

In the present study, the first experiment was carried out to investigate how observers make categorical decisions from morphed faces when only a limited amount of features is present. In this experiment, an upper face half, a lower face half, or an intact face was shown and observers were asked to categorize each morph as being either angry or fearful. The second experiment investigated how this relates to performance in a composite face task, when participants have to ignore one half of the face to make an optimal decision. In this experiment, a face assembled of an upper and lower half was presented. The observers had to judge only one half at a time, while the distractor half was showing either full-blown anger of fear. Together, these experiments aim to elucidate the psychophysics of recognizing facial expressions.

2. EXPERIMENT 1

2.1 Participants

30 participants took part in the experiment (22 female). Mean age of participants was 25 years (range: 18-32). Participants received course credit or 5 EUR for participation. Participants reported no history of neurological or psychiatric illness and had normal or corrected-to-normal vision. The study was approved by the ethics board of Bielefeld University (Ethic Statement Nr. 2014-010). All participants gave oral informed consent before taking part in the experiments.

Two participants showed performance close to guessing for all conditions combined with exceptionally fast reaction times, indicative of non-compliance. Their data were excluded, leaving 28 participants for further analysis.

2.2 Material

Anger and fear pictures of 20 identities (10 male, 10 female) were selected from the KDEF [23] and NimStim [24] databases. The pictures were used to generate morphs in 9 steps with Gimp 2.6 (www.gimp.org) and the GAP toolbox, resulting in 11 morphing grades including the original images. Faces were divided into an upper and lower half, with the border defined as being above the nostrils at the bridge of the nose of the middle morph. This border was defined for each face identity and subsequently applied to all of its morphs.

In Experiment 1 there were three visibility conditions (whole face, upper half, lower half). The half of no interest was blurred with a very broad Gaussian filter, preserving only a rough outline of the face, thereby conveying a feeling of wholeness, while rendering the masked features invisible (cf. Fig. 1).

[Figure 1. Design of Experiment 1. Illustration of a face morphed from the original fearful (outer left) to the original angry expression (outer right) in 9 intermediary steps, resulting in a total of 11 face morphs; a, whole face; b, upper half intact ('eyes' condition); c, lower half intact ('mouth' condition); due to copyright restrictions, the depicted example is an in-house generated face which was not used in the present experiment.]

2.3 Design

A two-alternatives forced choice identification task was used, in which participants had to decide for each face whether its expression was 'angry' or 'fearful'. Each of the 20 identities was presented in 11 morphing grades. The experiment consisted of two runs with a total of 40 trials per cell, resulting in a resolution of 2.5% for identification performance. The first run was repeated after a short break, to allow for analyses of intra-subject stability. This resulted in a total of 440 trials per masking condition (11 morphs x 20 identities x 2 runs). Pictures were shown with no time limit until a response was given, after which the next stimulus was shown directly, to allow for smooth work-flow. Participants were able to monitor their progress on a status bar presented at the bottom of the screen. Order of stimuli was randomized, the only constraint being that two subsequent trials never contained the same face identity.

Participants had to press the left or right mouse button to indicate whether the target face part showed an angry or fearful expression (button assignment counterbalanced between participants). Experiments were performed using Presentation software (Version 17.1, www.neurobs.com). All experiment files, including raw data, can be found in supplementary files S7 Code and S8 Dataset.

2.4 Analysis

To characterise the participants' performance, a logistic function was fitted to their data. The logistic function ($F_{logistic}(x;\alpha,\beta)=1/[\ 1+\exp(-\beta(x-\alpha))\ ]$) is well-suited to describe observers' performance in psychometric terms [25, p. 82]. It also provides slope information (β parameter) to quantify the steepness of the curve at the point of guessing. An s-shaped logistic function with a steep slope is indicative of high precision in distinguishing between two categories and little uncertainty [25,p. 20]. If categorical perception occurs, the steepest point in the psychometric curve corresponds to the category boundary, as this is the point where two stimuli will be discriminated best [5].

It should be noted that the logistic function might occur for any linear system limited by gaussian noise [26], including decisions based on simple image features such as contrast [27]. However, while the mere presence of a nonlinear or s-shaped curve in an identification task is not sufficient to claim the occurrence of categorical perception, its absence argues strongly against it. Therefore, a linear function ($F_{linear}(x;\alpha,\beta)=\beta*x+\alpha$), as an approximation to a logistic function with uniform slope was fitted to the data. If the logistic and linear function provide a similar fit for a certain condition, this would indicate that observers are unable or at least very poor at distinguishing between two distinct categories. Functions were fitted using an iterative optimization strategy, employing a non-linear least-squares procedure to derive the parameters for intercept and slope which describe the best-fitting function for the raw data. These parameters, together with the sum of squared errors (ss^2) of the best-fitting linear or logistic function were then compared between conditions.

Furthermore, for cross-validation, data from one half of the experiment were used to predict the other half of the data. Since each participant performed every trial twice, data from the first half of the experiment were used to predict the data from the second half (and vice versa). For each condition, one half of the participants' data was used to fit a logistic or linear function. Subsequently, the ss^2 between these data and the left-out half were computed. For example, a logistic function was fit to the first half of the whole face condition and this fitted function was then compared to the raw data of the second half of the experiment, using ss^2.

This allowed assessing whether it is possible to predict what condition a participant was in, given the responses made in a previous run and the function fitted to that data. The more similar two conditions are, the smaller the ss^2 should be. Using sums of squared errors has the advantages of being an intuitive measure of similarity and being virtually assumption-free [28, p. 106].

Data analysis was performed with Python 2.7.9 (www.python.org) using NumPy, SciPy, Pandas, Matplotlib, and the IPython Notebook, all as provided with Anaconda 2.2.0 (Continuum Analytics; docs.continuum.io/anaconda). Full code and output for all analyses can be found in the online supplement. Refer to S1 Code and S2 Code for data import and restructuring; S3 Code for main data analysis; S4 Code for fitting psychometric functions and cross-validation; and to S5 Code for analyses of single face identities. Analyses of Variance (ANOVA) for repeated measure designs were additionally carried out using IBM SPSS 22.

2.5 Results

2.5.1 Comparison of response curves for different masking types

Identification performance was compared across all 11 morphing grades for the three masking conditions (full face, upper face half, lower face half). Participants' performance is illustrated in Fig. 2. A 3x11 repeated-measures ANOVA showed a significant main effect of morphing grade, reflecting that anger responses increase as the faces are morphed towards this expression ($F_{(10,270)}=620.7$, $p<0.001$, $\eta_p^2=0.96$). There was no main effect of masking condition ($F_{(2,54)}=2.3$, $p=0.112$, $\eta_p^2=0.08$), indicating that there is no bias that would shift the responses between conditions. A significant condition-by-morphing grade interaction ($F_{(20,540)}=63.5$, $p<0.001$, $\eta_p^2=0.70$) indicated that the shape of the response curve differs between conditions. Paired t-tests for repeated measures and the Wilcoxon signed rank test as its non-parametric equivalent were carried out to test differences between conditions for all morphing grades. Results were only considered significant if the p-value for both metrics fell below $\alpha=0.05$.

Identification performance was best (i.e. significantly farthest from guessing) for the whole face, compared both to the upper (all *ps*<0.05, except 50% angry morphs) and lower half (all *ps*<0.05, except 60% angry morphs). Furthermore, performance was significantly better for the upper than for the lower half (all *ps*<.05 except for 50% and 60% angry morphs). This indicates that observers were able to make the most accurate decisions when presented with a whole face, followed by face halves containing only the eye region. Performance was worst for only the mouth region.

Judgements of the eye region closely resembled the whole face responses, as indicated by a significantly smaller difference between upper and whole face decisions compared to lower and whole face decisions (all *ps*<0.05 for pairwise comparisons of differences, except for 50% and 60% morphs).

[**Figure 2. Main Results Experiment 1.** Percent angry responses (y-axis) across the 11 morphing grades morphed from fear to anger (x-axis) for the three different visibility conditions (coloured lines); smaller right hand figures depict responses on single-participant level for each condition; 'whole', whole face is visible (no masking); ‘upper', upper face half is visible (lower half is masked); 'lower', lower face half is visible (upper half is masked).]

2.5.2 Cross-validation of conditions

If differences between conditions are genuine and stable, each condition should be able to predict itself best, e.g. the whole face condition of the first half of the experiment should be most similar to the whole face condition of the second half.

This was confirmed by split-half cross-validation with each half of the experiment used once as test and as training set, within each participant. Each condition was significantly closest to its analogue form the other half of the experiment, as expressed in significantly smallest ss^2 for these pairings (all $p<0.01$ for pairwise comparisons).

Furthermore, the upper and whole face conditions were significantly more similar to each other than to the lower half condition ($p<0.01$). This stronger similarity of the eye region to the whole face ratings further corroborates that the upper half carries more information about fear and anger expressions than does the mouth region.

2.5.3 Comparison of the superiority of a logistic fit to the data

A linear and logistic function were both fitted to the data of each participant and ss^2 were computed for all three masking conditions (Fig. 3a), with a better fit indicated by smaller ss^2. The superiority of the logistic over the linear function was significantly greatest for the whole face, followed by the upper face half, and lower face half, respectively ($F_{(2,54)}$=34.1, p<0.001, η_p^2=0.56; all ps<0.001 for pairwise comparisons).

[Figure 3. Curve Fitting for Experiment 1. Results of fitting different functions to the data set; red bars indicate 95% confidence intervals of the mean for each condition; line graphs in the background depict the raw data of each participant; black lines indicate that the participant's value for the left-hand condition is numerically bigger than the right-hand value; blue line indicates that the value is smaller for the left-hand compared to right-hand condition; a, superiority of fitting a logistic function to the data, with a higher difference of ss² compared to the linear function indicating a better fit; b, slope parameter of a logistic function fitted to the data; higher value indicates steeper slope; c, superiority of a logistic fit when the function is derived from one half of the experiment and ss² are computed to the left-out half ; 'whole', whole face is visible (no masking); 'upper', upper face half is visible (lower half is masked); 'lower', lower face half is visible (upper half is masked).]

2.5.4 Comparison of the slopes of the logistic function between conditions

When fitting a logistic function to each participant's data for each condition, the steepness of the slopes can be compared as measure for the presence of a category boundary, with a steep slope indicating stronger categorical perception (Fig. 3b).

Results show that the slopes were steepest for the whole face condition, followed by the upper and lower half ($F_{(2,54)}$=128.2, p<0.001, η_p^2=0.83; all ps<0.001 for pairwise comparisons), indicating a more abrupt shift from fear to anger responses when the whole face is visible and a smooth shift in responses when viewing the lower face half.

2.5.5 Cross-validation of response curves for different masking types

Left-out data from one half of the experiment were predicted using data from the other half of the experiment with the two runs of the experiment used for cross-validation. Each condition was modelled for one half of the experiment using a linear or logistic function fitted to these data. These training sets were then compared to the left-out half of the experiment to check how the models compare to each other in predicting new data.

The results of the cross-validation analysis (Fig. 3c) show that the superiority of a logistic over a linear function for predicting left-out data is largest for the whole face condition and weakest when only the lower face half is visible ($F_{(2,54)}$=21.0, p<0.001, $\eta_p^2=0.44$; ps<0.05 for all pairwise comparisons).

3. EXPERIMENT 2

3.1 Methods

Mean age of participants was 24 years (range 19-29). Of the 30 participants, 24 were female. Participants received course credit or 7EUR for participation and gave oral informed consent before starting the experiment.

The same boundary to divide faces into lower and upper halves as in experiment 1 was used and all other design choices were identical to experiment 1, with the following exceptions: This time 100% fearful or 100% angry expression halves were combined with the morphed pictures to create the composite face illusion. There were four conditions: 'lower half 100% angry' (Fig. 4a), 'lower half 100% fearful (Fig. 4b), 'upper half 100% angry' (Fig. 4c) and 'upper half 100% fearful' (Fig. 4d). The complementary half was presented in 11 morphing steps, and was the part which the observers were asked to identify as either angry of fearful. This part was always framed by a red square (cf. Fig. 4). Participants were explicitly instructed to focus only on the framed face half and to ignore the other half. Unlike experiment 1, order of conditions was blocked, to make it easier for the participants to focus on the part of interest. Each block required either always judging the upper half or always judging the lower half, and consisted of 440 trials, with block order counterbalanced between participants. Within blocks, order of the 20 face identities, the 11 morphed face halves and whether the to-be-ignored half was fearful or angry were randomised.

[Figure 4. Design of Experiment 2 (Composite Face). Illustration of a face morphed from the original fearful (outer left) to the original angry expression (outer right) in 9 intermediary steps, resulting in a total of 11 face morphs; a, eye judgements with 100% angry lower face half; b, eye judgements with 100% fearful lower half; c, mouth judgements with 100% angry upper face half; d, mouth judgements with 100% fearful upper half; conditions a and b or conditions c and d were always presented in one block, to aid participants in focusing on one face half only; due to copyright restrictions, the depicted example is an in-house generated face which was not used in the present experiment.]

One participant reported to have only focused on the mouth region throughout the experiment, ignoring which face half was currently framed. These data were excluded, leaving 29 data sets for analysis.

3.2 Results

3.2.1 Comparison of response curves for different composite faces

Depending on the viewing condition, responses for the attended face half were biased in the direction of the to-be-ignored face half (Fig. 5a).

A 2x2x11 ANOVA with the factors upper/lower face half, anger/fear distractor and morphing grade (fear to anger) revealed a main effect of morphing grade ($F_{(10,280)}$=213.7, p<0.001, $\eta_p{}^2$=0.88), a main effect of emotion ($F_{(1,28)}$=109.6, p<0.001, $\eta_p{}^2$=0.80) and significant interactions between all factors (all ps<0.001). The main effect of emotion in the distractor half reflects that responses for the eye region were shifted towards anger when the lower face half was angry, and towards fear when the lower half was fearful (p<0.01 for all 11 morphing grades). The same bias occurred for the judgements of the mouth region, when the upper half was either angry or fearful (Fig. 5c; all p<0.001).

The interaction effect between emotion and face half ($F_{(1,28)}$=26.0, p<0.001, $\eta_p{}^2$=0.48) reflects that the bias was significantly stronger for mouth judgements than for eye judgements, indicating that the upper face half exerts a stronger influence on the lower half than vice versa. Across all 11 morphing grades, fearful and angry distracting halves led to bigger differences in mouth judgements, compared to the differences in eye judgements (all $p<0.001$, for pairwise comparisons of differences).

[Figure 5. Main Results for Composite Faces. Percent angry responses (y-axis) across the 11 morphing grades morphed from fear to anger (x-axis) for the four different composite face conditions (coloured lines); left-hand figures depict mean results with 95% confidence intervals on group level; right-hand figures depict responses on single-participant level for each condition; 'lower half fearful', observers judge the upper half, while the lower half is always 100% fearful; 'lower half angry', observers judge the upper half, while the lower half is always 100% angry; 'upper half fearful', observers judge the lower half, while the upper half is always 100% fearful; 'upper half angry', observers judge the lower half, while the upper half is always 100% angry.]

3.2.2 Cross-validation of conditions

Split-half cross-validation with each half of the experiment used once as test and as training set showed that each condition was significantly closest to its analogue form the other half of the experiment (all $p<0.01$ for pairwise comparisons), indicative of reliable differences between conditions.

3.2.3 Comparison of the superiority of a logistic fit to the data

The superiority of a logistic function over a linear function was significantly stronger for eye judgements as compared with mouth judgements (Fig 6a; main effect of face half $F_{(1,28)}=13.7$, $p<0.001$, $\eta_p^2=0.33$; all $p<0.05$ for pairwise comparisons), indicating that when observers judge the upper face half, their response function is significantly more s-shaped that when they judge the lower face half.

[**Figure 6. Curve Fitting for composite faces.** Results of fitting different functions to the data set; red bars indicate 95% confidence intervals of the mean for each condition; line graphs in the background depict the raw data of each participant; 'rating upper/lower fearful', observers judge the upper half, while the lower half is always 100% fearful; 'rating upper/lower angry', observers judge the upper half, while the lower half is always 100% angry; 'rating lower/upper fearful', observers judge the lower half, while the upper half is always 100% fearful; 'rating lower/upper angry', observers judge the lower half, while the upper half is always 100% angry; for more information please refer to the legend of figure 3.]

3.2.4 Comparison of the slopes of the logistic function between conditions

When fitting a logistic function to each participant's data for each of the three conditions, the steepness of the slopes was significantly higher for upper than for lower face half judgements (Fig. 6b; main effect of face half; $F_{(1,28)}$=81.4, p<0.001, η_p^2=0.74; all ps<0.001 for pairwise comparisons), indicative of a more abrupt change in fear/anger decisions when viewing the upper half. However, since lower half judgements were biased by the eye region to the point where for many participants performance never crossed 50% guessing, slope parameters for the mouth regions are difficult to interpret.

3.2.5 Cross-validation of response curves for different masking types

Analogous to experiment 1, results of the cross-validation analysis show that the superiority of a logistic over a linear function for predicting left-out data is larger for upper half versus lower half judgements (Fig. 6c), but only significant when the ignored face half was fearful (p<0.05). The difference was only numerically superior when an anger mouth was present (significant interaction effect $F_{(1,28)}$=4.8, p=0.038, η_p^2=0.15).

4. RESULTS FOR SINGLE FACE IDENTITIES

Since 20 different face identities were used as stimuli in the experiments, the present results may underestimate or blur the sigmoid shape of the response functions, due to averaging over a heterogeneous set of stimuli with shifted intercepts.

Therefore, data of both experiments were re-analysed by fitting a logistic function to the responses to each of the 20 individual face identities, averaged across participants. Subsequently, the intercepts were shifted to be uniformly at 50%, so that each curve is at 50% guessing (y-axis) when the face is 50% angry (x-axis). The necessary shift was determined from the fitted logistic functions and then applied to the underlying raw data, allowing to compare how the curves would be shaped, given equal intercepts for all face identities. The results are illustrated in Fig. 7 and are described in-depth in S5 Code. Visual inspection of the data suggests that a sigmoid function clearly emerges only for whole faces and upper halves in experiment 1 and for upper halves in experiment 2 when accounting for potential variability between faces.

[Figure 7. correction for different intercepts between face identities. Main analyses for both experiments after modelling responses to each face identity individually and shifting the intercepts to be uniformly at 50%; a, results for experiment 1 (masking); b, results for experiment 2 (composite faces); note that a different number of data points is available for each bin on the x-axis, and therefore estimates at the extremes are less reliable; especially in experiment 2, mouth judgements were strongly biased, with few responses at 50% guessing; please refer to Supplement S5 for more details on the analysis procedure.]

5. DISCUSSION

The present study set out to investigate what part of the face is most diagnostic for categorical perception of fear-anger morphed expressions, and how this information is integrated when presented in a composite face.

Experiment 1 showed that the upper face half is more diagnostic than the lower half for deciding between fearful and angry expressions in an identification task. The upper half alone was sufficient for high identification performance, closely reflecting the whole face condition. Its perception could also be well-described by a logistic function with a steep category boundary, providing converging evidence for categorical perception based on the upper face half alone. When viewing the lower half, responses were significantly closer to chance, were less s-shaped and had a comparably shallower slope.

In experiment 2, it was expected that the to-be-ignored halves of the face would bias performance, and that regions that have proven most diagnostic in experiment 1 should exert a stronger biasing influence over complementary face parts than vice versa. Accordingly, results indicated that both eye and mouth regions bias judgements on the respective other face half, with strongest influences exerted by the eyes, leading to what might be considered a breakdown of performance for mouth judgements. This effect is noteworthy, as information in the mouth region is equally present in both experiments. Nevertheless, when instructed to evaluate the mouth and ignore the eyes in a composite face, the participants' ratings reflect the expression of the eyes numerically more strongly than they reflect mouth judgements. This means that a cue that is reasonably useful for expression recognition in isolation, will be of almost no use when exposed to the influence of a strongly diagnostic cue.

The influence of the mouth on eye judgements should be considered equally striking, since it illustrates that even a very diagnostic cue that allows for recognition levels almost as high as a whole face, can be subject to significant bias form a substantially weaker cue. This suggests that even when analytical processing of a single feature is sufficient to allow for categorical perception, human observers seem to integrate features into a whole whenever possible.

The dominance of the eye region in the present experiments is well in line with masking studies which point towards its prominent role for both fear and anger recognition [13]. Studies with neurological patients also indicate a specific importance of the eyes in fear detection [22]. Also, when asked to evaluate expressions of emotion, observers most readily fixate the eyes [14], irrespective of the emotion in question.

Regarding identity recognition, there are many studies that demonstrate the importance of the eyes. For example, when a face has to be learned in a training run and later distinguished from distractors differing in one feature, it is easier to recognize identity changes due to differences in the eyes, compared to nose or mouth changes [29]. Masking out the eyes may even represent a legal requirement in some contexts for ensuring anonymity. Although the scientific basis for the effectiveness of this procedure is equivocal at best [cf. 30], it illustrates the paramount importance we intuitively place on the eye region. Regarding mental states, it is also commonly assumed that they can be assessed best from the eye region, as reflected for example in the 'reading the mind in the eyes' test [31]. Accordingly, one must ask whether the eye advantage reported in the present study can be generalised to other features and other expressions of emotion.

Even though the eyes may be most likely fixated when judging emotions from intact faces [14], this does not necessarily imply that most information is coded in the eye region. This becomes obvious when considering the studies carried out with patient SM, who exhibits major difficulties at recognizing fear from faces, due to a failure to fixate the eyes [22]. Nevertheless, she is virtually unimpaired at recognizing all other basic emotions. This indicates that it must be possible to recognize those emotions utilising other diagnostic features. Accordingly, masking studies [13] and coding systems for basic emotion expressions [10], suggest that each expression can be best recognized from a set of features specific to that emotion and that focusing on the eyes cannot be the optimal strategy in all cases.

Therefore, regarding the results of the present experiments, we would predict that the basic mechanisms outlined here will be replicated with other facial expressions, but perhaps not for the same features. For example, happy-surprised pairings should be best recognized from the mouth, as both involve mainly muscles of the lower face half [10,13,32]. Interestingly, eye-tracking studies found that more fixations on the eye regions correlates with improved emotion recognition for all expressions [14,33] and argue that fixating the eyes reflects a holistic processing strategy. One could test this assumption using happy-surprise pairings and the composite face task from experiment 2. If holistic processing is associated with fixating the eyes, then eye judgements should be easier in a composite face task, even if a strongly diagnostic distracting lower face half is present.

The present findings should not only be extended to investigate other expressions, but also further elaborated for the currently used fear-anger pairing, to shed light on the basic perceptual mechanisms at work. The employed identification task could be complemented by a discrimination task. There, it should be easier to discriminate two images when they are on opposing sides of the category boundary, compared with being both sampled from within the same category [5,6]. Thus, the sigmoid curves from the present experiments could be used to predict the discrimination performance in a follow-up experiment. As differences in slope between conditions can reflect differences in variance of the underlying noise distribution or the uncertainty of the observer [34,35], it would be important to establish that a between-category advantage [5] is indeed present at the point where the psychometric function is steepest in the identification task.

On the basis of the present data from experiment 1, we would therefore predict that discrimination performance at the category boundary should be higher for eye than mouth judgements.

In the composite face task, the peak of the discrimination function should be shifted to the right and left, respectively, when judging the eyes, in accordance with the intercept shifts due to the mouth expressions. Finally the discrimination function should be almost flat when judging the mouth in a composite face, as there is virtually no crossing of a category boundary in the present data.

To further generalise the present findings regarding fear-anger pairings, averaged faces could be used, which have the advantage that idiosyncrasies would average out while expressions prototypical for a certain emotion would be emphasized [36, p. 76]. While studies on categorical perception are often performed with a very limited number of identities (<=4; [cf. 4,7,8,21]), the 20 different faces used in the present experiments should be considered a valuable step towards greater generalisability. However, the fact that the mouth provided so little diagnostic information might be partly due to the fact that 12 out of the 20 face identities expressed the emotions with a closed mouth. While this might have reduced variance in the expressiveness of the lower face half, both the pressing of the lips as well as the baring of the teeth are valid anger expressions [10,12]. Of the two, lip pressing is probably much more frequent in everyday life. Also, when making statements about the importance of either eyes or mouth for recognizing an expression, it should be kept in mind that these features were rather crudely operationalised as the upper and lower face half, respectively. While such a division is standard practice in the field [18,21,37], the muscles around the nose have likely contributed to the judgements of both face halves, and the eyebrows [38] and forehead [11] might have contributed to judgements of the upper face half. While the sclera of the eyes seems crucial for fear detection [39], anger judgements might depend more on the surrounding muscles of the eyes and the eyebrows [13,38], especially since patient SM shows no difficulty at recognizing anger [22].

Another interesting question is to what degree the failure to concentrate on one half of a composite face can be voluntarily overridden, especially since previous work has shown that emotion recognition improvements are associated with increased fixation of the eyes [33]. This could be investigated by introducing training blocks before the experiment proper, a forced minimum exposure time to discourage spontaneous responses, biofeedback from an eye-tracking device helping to fixate on a specific face part, and rewards for correct responses.

To summarise, the present studies show that identification of fear and anger in morphed faces relies heavily on the eye region, corroborating previous research [10,13,14,19]. Expanding previous work, the study shows that observers can perform categorical perception even when viewing only a single face part. For the employed fear-anger pairings, the eye region was sufficient to allow for decisions with a high degree of certainty. This demonstrates that categorical perception can emerge based on single features only, adding to our understanding of how categorization works in complex stimuli such as faces.

Even though the eye region was almost as diagnostic as the whole face, it was not immune to biasing influences in a composite face task. When a full face with conflicting expressions was shown, the psychometric curves for eye judgements were shifted in accordance with the expression in the lower half of the face. For the first time, we can therefore show that holistic perception of a full face can be explained based on the diagnostic value of its constituent parts. In addition, given that observers were aware that a holistic processing strategy impaired their performance, this illustrates that humans involuntarily process faces as a gestalt when trying to make categorical decisions about its expression.

ACKNOWLEDGEMENTS

We would like to thank Maleen Fiddicke for help with data collection.

REFERENCES

1. Keltner D, Haidt J. Social Functions of Emotions at Four Levels of Analysis. Cogn Emot. 1999;13: 505–521. doi:10.1080/026999399379168

2. Ekman P. Are there basic emotions? Psychol Rev. 1992;99: 550–553. doi:10.1037/0033-295X.99.3.550

3. Du S, Tao Y, Martinez AM. Compound facial expressions of emotion. Proc Natl Acad Sci. 2014;111: E1454–E1462. doi:10.1073/pnas.1322355111

4. Young AW, Rowland D, Calder AJ, Etcoff NL, Seth A, Perrett DI. Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. Cognition. 1997;63: 271–313. doi:10.1016/S0010-0277(97)00003-6

5. Fugate JMB. Categorical Perception for Emotional Faces. Emot Rev. 2013;5: 84–89. doi:10.1177/1754073912451350

6. Harnad SR. Categorical Perception. Categorical perception: the groundwork of cognition. Cambridge; New York: Cambridge University Press; 1990.

7. Calder AJ, Young AW, Perrett DI, Etcoff NL, Rowland D. Categorical Perception of Morphed Facial Expressions. Vis Cogn. 1996;3: 81–118. doi:10.1080/713756735

8. De Gelder B, Teunisse J-P, Benson PJ. Categorical Perception of Facial Expressions: Categories and their Internal Structure. Cogn Emot. 1997;11: 1–23. doi:10.1080/026999397380005

9. Etcoff NL, Magee JJ. Categorical perception of facial expressions. Cognition. 1992;44: 227–240. doi:10.1016/0010-0277(92)90002-Y

10. Ekman P, Friesen WV, Hager JC. Facial action coding system. Salt Lake City UT Hum Face. 2002;

11. Ekman P, Friesen WV. Unmasking the face: a guide to recognizing emotions from facial clues. Cambridge, MA: Malor Books; 1975.

12. Matsumoto D, Ekman P. Facial expression analysis. Scholarpedia. 2008;3: 4237. doi:10.4249/scholarpedia.4237

13. Smith ML, Cottrell GW, Gosselin F, Schyns PG. Transmitting and Decoding Facial Expressions. Psychol Sci. 2005;16: 184–189. doi:10.1111/j.0956-7976.2005.00801.x

14. Guo K. Holistic Gaze Strategy to Categorize Facial Expression of Varying Intensities. Verdejo García A, editor. PLoS ONE. 2012;7: e42585. doi:10.1371/journal.pone.0042585

15. Rossion B. Picture-plane inversion leads to qualitative changes of face perception. Acta Psychol (Amst). 2008;128: 274–289. doi:10.1016/j.actpsy.2008.02.003

16. Sergent J. An investigation into component and configural processes underlying face perception. Br J Psychol. 1984;75: 221–242. doi:10.1111/j.2044-8295.1984.tb01895.x

17. McKone E, Yovel G. A single holistic representation of spacing and feature shape in faces. J Vis. 2010;8: 163–163. doi:10.1167/8.6.163

18. Young AW, Hellawell D, Hay DC. Configurational information in face perception. Perception. 1987;16: 747–759. doi:10.1068/p160747

19. Tanaka JW, Kaiser MD, Butler S, Le Grand R. Mixed emotions: Holistic and analytic perception of facial expressions. Cogn Emot. 2012;26: 961–977. doi:10.1080/02699931.2011.630933

20. Calder AJ, Jansen J. Configural coding of facial expressions: The impact of inversion and photographic negative. Vis Cogn. 2005;12: 495–518. doi:10.1080/13506280444000418

21. Chen M-Y, Chen C-C. The contribution of the upper and lower face in happy and sad facial expression classification. Vision Res. 2010;50: 1814–1823. doi:10.1016/j.visres.2010.06.002

22. Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. Nature. 2005;433: 68–72. doi:10.1038/nature03086

23. Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces (KDEF). CD ROM Dep Clin Neurosci Psychol Sect Karolinska Institutet. 1998; 91–630.

24. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: Judgments from untrained research participants. Psychiatry Res. 2009;168: 242–249. doi:10.1016/j.psychres.2008.05.006

25. Kingdom FAA, Prins N. Psychophysics: a practical introduction. London: Academic; 2010.

26. Duda RO, Hart PE, Stork DG. Pattern classification [Internet]. 2001. Available: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=531488

27. Foley JM, Legge GE. Contrast detection and near-threshold discrimination in human vision. Vision Res. 1981;21: 1041–1053. doi:10.1016/0042-6989(81)90009-2

28. Hilborn R. The ecological detective: confronting models with data. Princeton, NJ: Princeton University Press; 1997.

29. Tanaka JW, Farah MJ. Parts and wholes in face recognition. Q J Exp Psychol Sect A. 1993;46: 225–245. doi:10.1080/14640749308401045

30. Newton EM, Sweeney L, Malin B. Preserving privacy by de-identifying face images. IEEE Trans Knowl Data Eng. 2005;17: 232–243. doi:10.1109/TKDE.2005.32

31. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. J Child Psychol Psychiatry. 2001;42: 241–251. doi:10.1111/1469-7610.00715

32. Leppänen JM, Hietanen JK. Is there more in a happy face than just a big smile? Vis Cogn. 2007;15: 468–490. doi:10.1080/13506280600765333

33. Pollux PMJ, Hall S, Guo K. Facial Expression Training Optimises Viewing Strategy in Children and Adults. Barton JJS, editor. PLoS ONE. 2014;9: e105418. doi:10.1371/journal.pone.0105418

34. Ashby FG, Townsend JT. Varieties of perceptual independence. Psychol Rev. 1986;93: 154–179.

35. Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. J Opt Soc Am A. 1985;2: 1508–1532.

36. Bruce V, Young AW. Face perception. London ; New York: Psychology Press; 2012.

37. Grand RL, Mondloch CJ, Maurer D, Brent HP. Impairment in Holistic Face Processing Following Early Visual Deprivation. Psychol Sci. 2004;15: 762–768. doi:10.1111/j.0956-7976.2004.00753.x

38. Tipples J, Atkinson AP, Young AW. The eyebrow frown: A salient social signal. Emotion. 2002;2: 288–296. doi:10.1037/1528-3542.2.3.288

39. Whalen PJ. Human Amygdala Responsivity to Masked Fearful Eye Whites. Science. 2004;306: 2061–2061. doi:10.1126/science.1103617

SUPPLEMENT

S1 Code. Data import for experiment 1

S2 Code. Data import for experiment 2

S3 Code. Main analysis and plotting

S4 Code. Curve fitting and cross-validation

S5 Code. Additional analyses for face identities

S6 Code. Complete analysis code from S1-S5 in executable IPython Notebook format

S7 Code. Complete code for experiment presentation in Neurobs Presentation format

S8 Dataset. Complete raw data of all participants

Fig1

Click here to download Figure: Fig1.tiff

Fig2

Click here to download Figure: Fig2.tiff

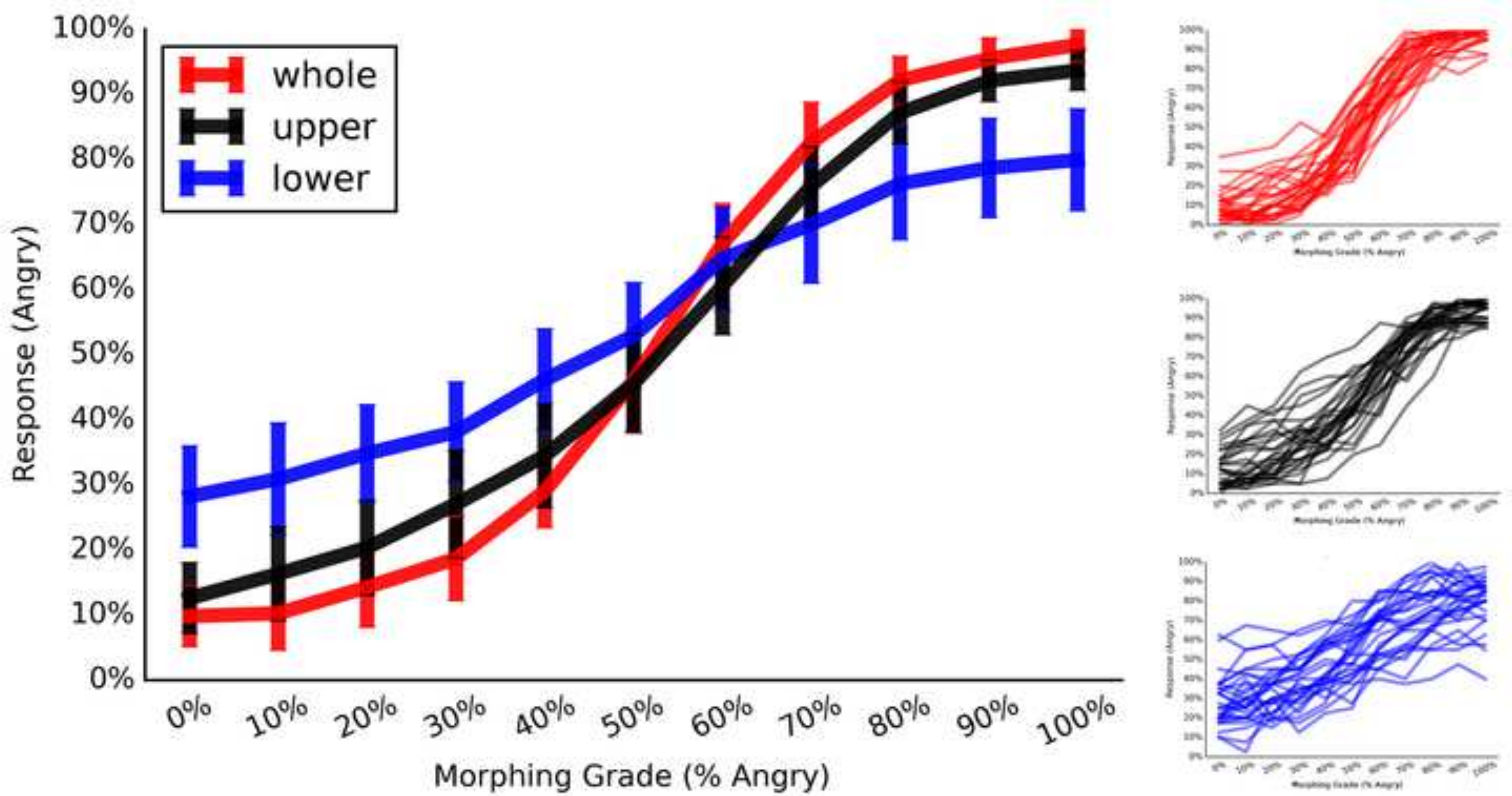


Fig3
Click here to download Figure: Fig3.tiff

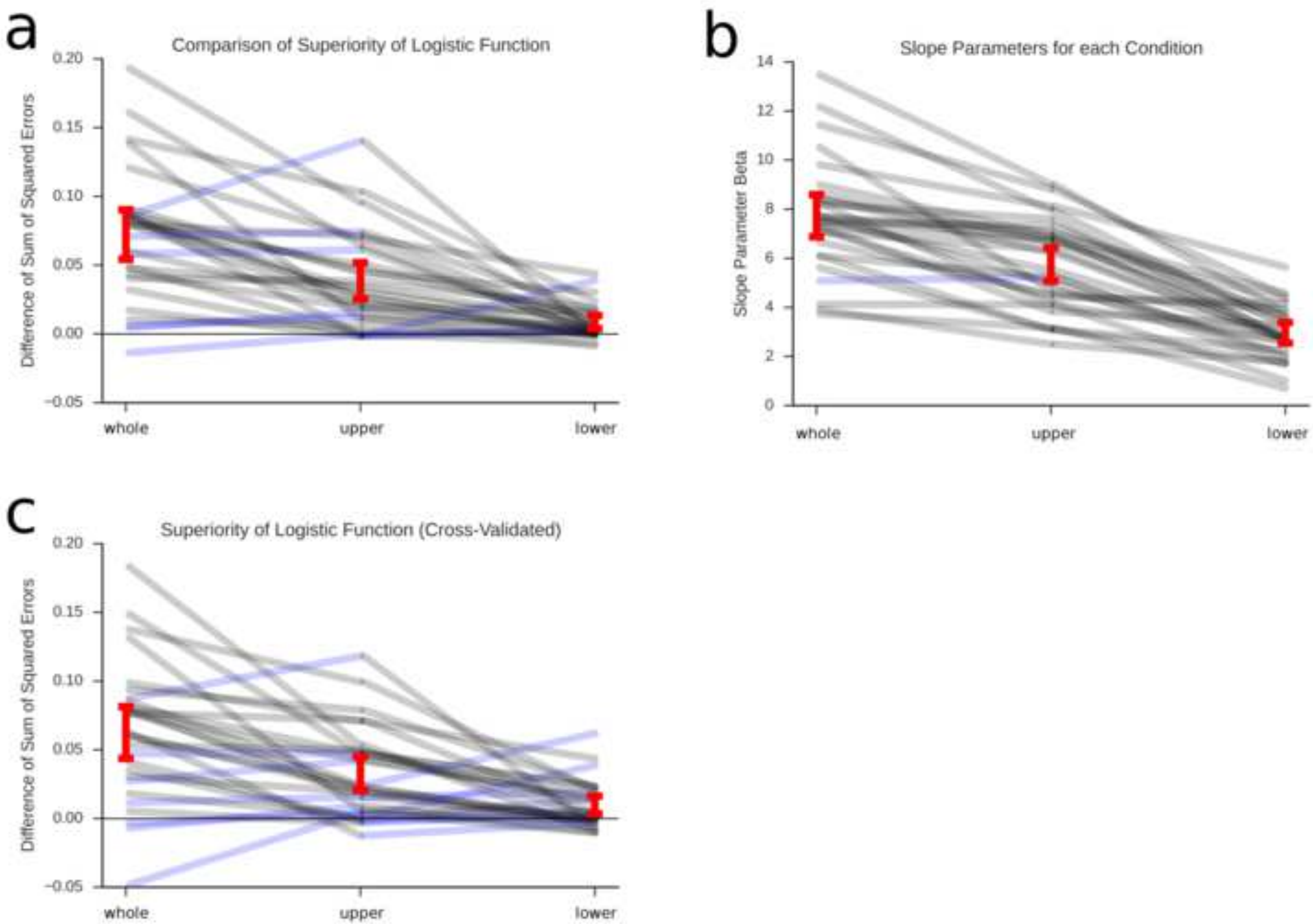


Fig4
Click here to download Figure: Fig4.tiff

Fig5

Click here to download Figure: Fig5.tiff

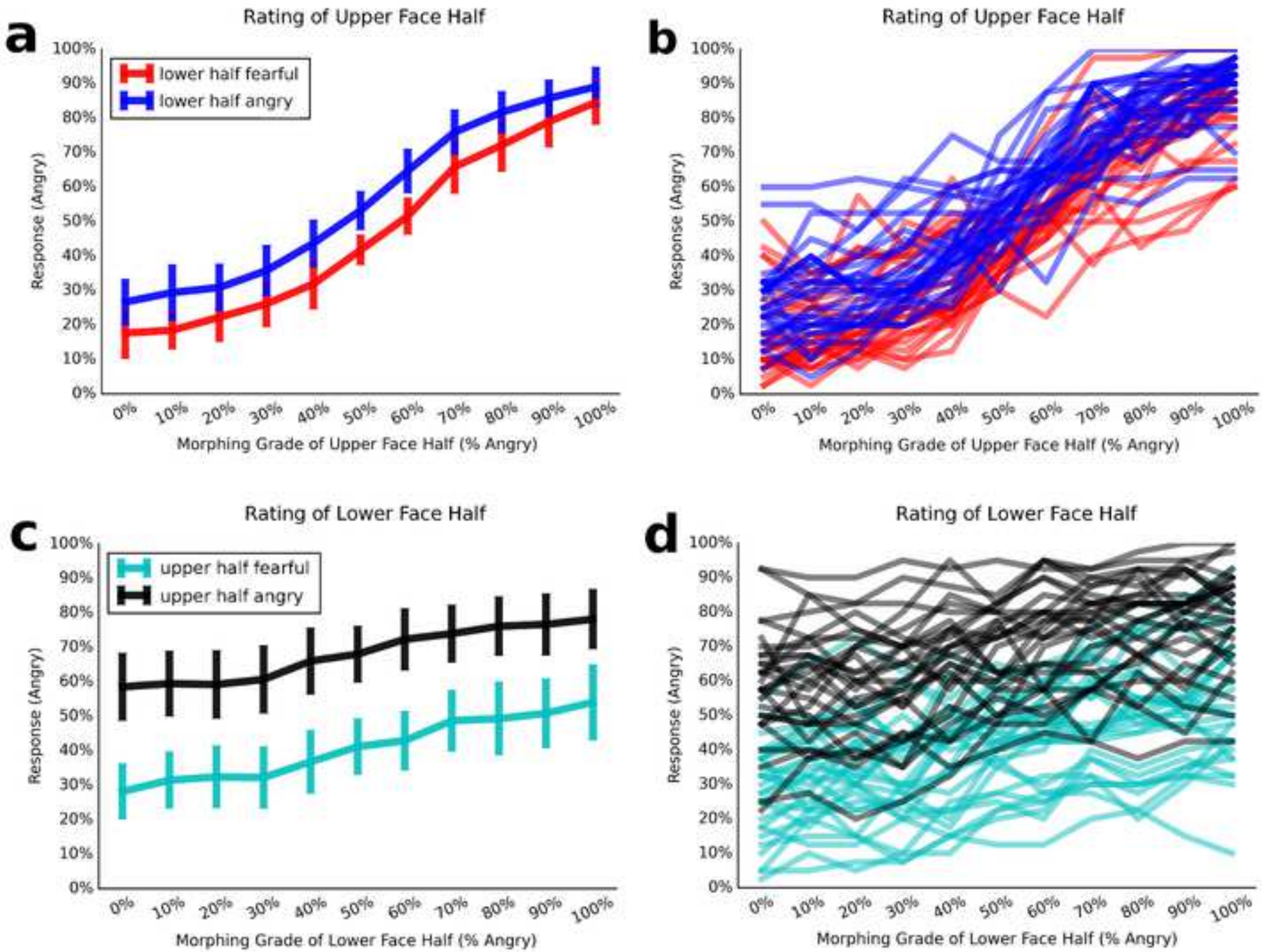


Fig6

Click here to download Figure: Fig6.tiff

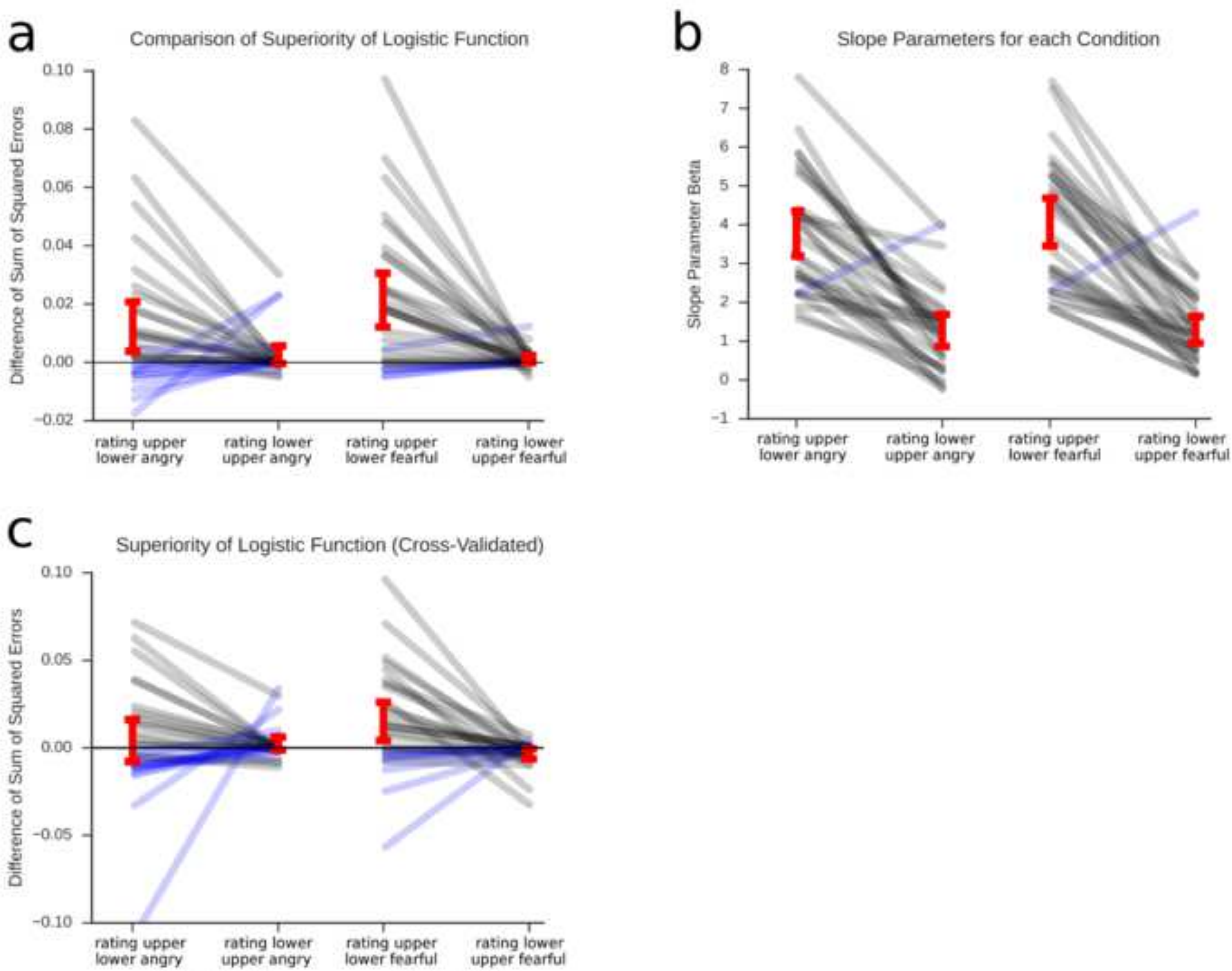
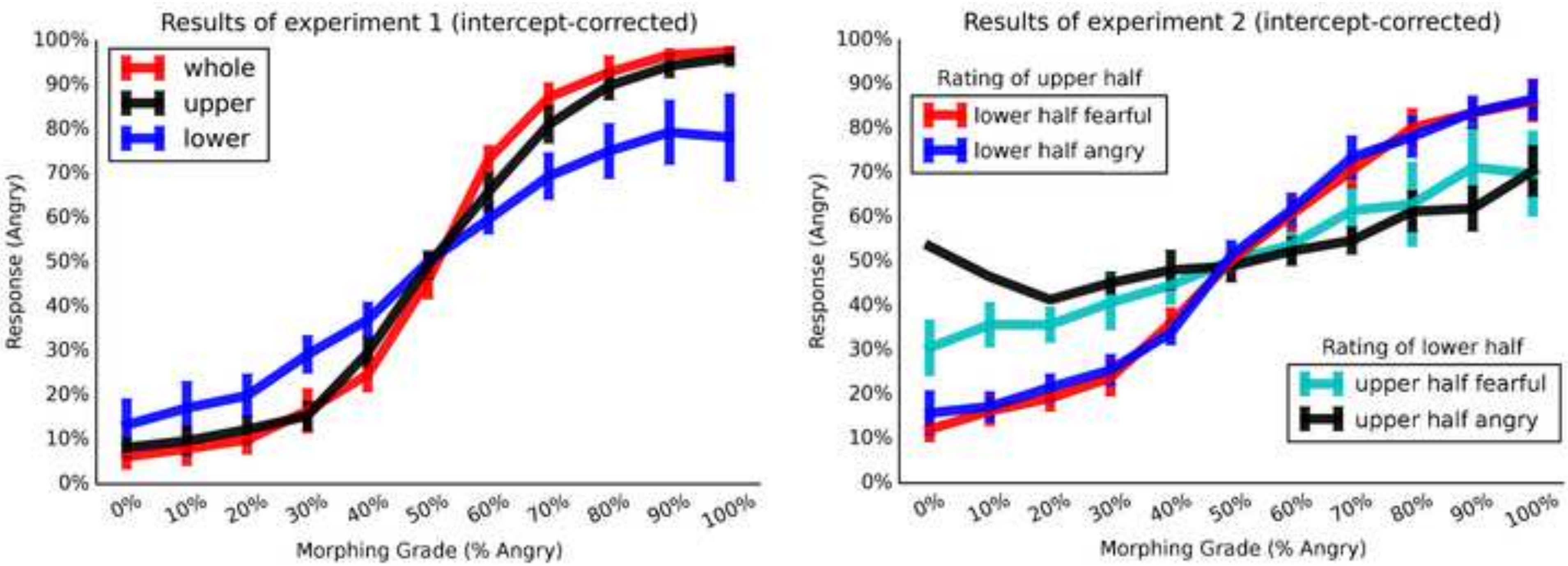


Fig7

Click here to download Figure: Fig7.tiff



Click here to download Supporting Information: S1_Code.html

Click here to download Supporting Information: S2_Code.html

Click here to download Supporting Information: S5_Code.html

Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces.

Martin Wegrzyn[1,2,*], Isabelle Bruckhaus[1] & Johanna Kissler[1,2]

[1]Department of Psychology, Bielefeld University, Bielefeld, Germany

[2]Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Bielefeld, Germany

Manuscript Requirements:

Title: 90 Characters

Abstract: 283 Words

Text: 5498 Words

References: 39

Tables: 0

Figures: 7

***Corresponding Author:**

Martin Wegrzyn

Bielefeld University

Department of Psychology

Postfach 10 01 31

33501 Bielefeld

Tel. +49 521-106-4527

Fax. +49 521-106-154454

martin.wegrzyn@uni-bielefeld.de

ABSTRACT

Human observers are remarkably proficient at recognizing expressions of emotions and at readily grouping them into distinct categories. When morphing one facial expression into another, the linear changes in low-level features are insufficient to describe the changes in perception, which instead follow an s-shaped function.

Important questions are, whether there are single diagnostic regions in the face that drive categorical perception for certain parings of emotion expressions, and how information in those regions interacts when presented together.

We report results from two experiments with morphed fear-anger expressions, where (a) half of the face was masked or (b) composite faces made up of different expressions were presented.

When isolated upper and lower halves of faces were shown, the eyes were found to be almost as diagnostic as the whole face, with preserved categorical perception. In contrast, the mouth allowed for a substantially lesser amount of accuracy and nearly absent categorical perception.

When a composite face consisting of mismatched upper and lower halves was used and observers were instructed to exclusively judge either the expression of mouth or eyes, the to-be-ignored part always influenced perception of the target region. In line with experiment 1, the eye region exerted a much stronger influence on mouth judgements than vice versa. Again, categorical perception was significantly more pronounced for upper halves of faces.

The present study shows that identification of fear and anger in morphed faces relies heavily on information from the upper half of the face, particularly the eye region. Categorical perception is possible when only the upper face half is present, but

compromised when only the lower part is shown. Moreover, observers tend to integrate all available features of a face, even when trying to focus on only one part.

1. INTRODUCTION

Facial expressions are a powerful means of conveying information about the emotional state of an individual. Recognizing and correctly interpreting these non-verbal signs is of vital importance for successful social interaction [1].

Faces are not only complex stimuli, but their expressions can often be subtle and ambiguous. Despite this, human observers can readily recognize expressions of emotions and assign them distinct labels. A large body of research suggests that there are a number of basic emotions which are correctly recognized by most observers [2], although their exact number is under debate [3]. Even when the expressions' intensity varies continuously, as in manipulated faces where one expression is gradually morphed into another, human observers readily group them into distinct categories with a high level of confidence [4].

Such categorical perception is demonstrated by identification tasks where linear changes in low-level features (e.g. basic visual properties like colour or shape) lead to non-linear changes in perception, which are best described by a sigmoid function with a steep category boundary. In general, categorical perception refers to the phenomenon that objects from the same category will be perceived as more similar than objects from different categories [5], despite being equally far apart from each other on a given physical dimension. One example is human colour perception, where two different wavelengths will only be perceived as two different colours when a category boundary is crossed [6]. For facial expressions, categorical perception has been demonstrated in a number of seminal studies [7–9].

Faces are inherently multidimensional stimuli, and in face processing categorical perception can relate to different dimensions (identity, age, attractiveness, gender, emotion), which in turn may depend on different facial features. Regarding emotion, each

expression of a basic emotion can be described in terms of a number of muscle groups that are active when a person shows that particular emotion [10]. For example, fear will most frequently be expressed in the face by the brows raised and drawn together, the upper eyelids raised, the lower eyelids tensed, the lips stretched back and, in some cases, the mouth opened [11,12]. Furthermore, masking studies have demonstrated that each basic emotion is recognized through different diagnostic areas of the face [13]. For example, fear can be best recognized when the eyes are visible, while happiness is best inferred from the mouth region. On the other hand, eye tracking studies have shown that for all six basic expressions, including ones of low intensity, people focus primarily on the eyes [14].
While eye tracking studies shed light on how observers usually inspect faces under free viewing conditions, masking studies help to understand what information is indispensable in order to make a correct decision. Therefore a masking approach helps to make causal inferences about the diagnostic value of each facial feature.
To understand categorical perception of facial expressions, it is essential to know which information in the face is used to make a categorical decision and how facial features are integrated into a perceptive whole, so that successful categorization can occur.
The term 'holistic processing' refers to the fact that faces are preferentially processed as a whole [15–17]. Findings from the composite face illusion, where complementary upper and lower face halves are combined into a whole face, serve to illustrate this mechanism: It is considerably more difficult to make a decision about the properties of either half in such composites than it is to judge each part in isolation [18]. It may be more difficult to recognize a person's identity from a lower face half when it has been combined with the upper face half of another person, while the same task is easy when each half is presented in isolation [18]. Using happy and angry faces, interference from composites with mis-matching expressions have been found reflected in slower reaction times [19,20].

So far, both masking studies [13] and composite face paradigms [19] used full-blown facial expressions instead of gradually morphed faces. Chen and Chen [21] have used morphed happy-sad expressions in a composite face, but always asked to judge the whole face and not its parts, therefore making it difficult to draw inferences on how the perception of one facial feature is influenced by another one.

Masking or composite face studies with morphed faces however allow to better describe human behaviour in psychophysical terms, as the subjective perception dissociates from the changes in low-level features. They also allow to better understand how we process subtle and ambiguous expressions, which are arguably most relevant in daily life. Therefore, it is useful to investigate the relationship of categorical perception and featural (i.e. focusing on single parts) or configural (i.e. integrating parts into a whole) face processing with masked and composite faces. Main questions are, whether there are diagnostic regions in the face that are sufficient to allow for categorical perception, and whether multiple features need to be integrated when making categorical judgements. Therefore, to better understand mechanisms of categorical perception of facial expressions, a promising approach is to break down the face into its features and understand how each contributes to identifying a certain expression. In a second step, the features can be re-assembled into a whole and one can investigate how they are integrated to infer an expression from a whole face.

To address these questions, two experiments with faces morphed from fear to anger were performed. Beginning with the seminal work by Etcoff and Magee [9], fear-anger pairings have been frequently used for investigating categorical perception (e.g. [7,8] ; cf. [5]). Previous masking studies with full-blown fear and anger expressions have shown that both expressions are mainly recognized from the eye region [13] and studies with neurological patients also point to a special role of the eyes for recognizing fear [22]. Therefore, one can hypothesize that categorical perception of fear and anger will rely mostly on the features in the upper half of the face. In line with research on the composite face illusion [19,20], the more informative half of the face should also dominate the perception of a full face.

In the present study, the first experiment was carried out to investigate how observers make categorical decisions from morphed faces when only a limited amount of features is present. In this experiment, an upper face half, a lower face half, or an intact face was shown and observers were asked to categorize each morph as being either angry or fearful. The second experiment investigated how this relates to performance in a composite face task, when participants have to ignore one half of the face to make an optimal decision. In this experiment, a face assembled of an upper and lower half was presented. The observers had to judge only one half at a time, while the distractor half was showing either full-blown anger of fear. Together, these experiments aim to elucidate the psychophysics of recognizing facial expressions.

2. EXPERIMENT 1

2.1 Participants

30 participants took part in the experiment (22 female). Mean age of participants was 25 years (range: 18-32). Participants received course credit or 5 EUR for participation. Participants reported no history of neurological or psychiatric illness and had normal or corrected-to-normal vision. The study was approved by the ethics board of Bielefeld University (Ethic Statement Nr. 2014-010). All participants gave oral informed consent before taking part in the experiments.

Two participants showed performance close to guessing for all conditions combined with exceptionally fast reaction times, indicative of non-compliance. Their data were excluded, leaving 28 participants for further analysis.

2.2 Material

Anger and fear pictures of 20 identities (10 male, 10 female) were selected from the KDEF [23] and NimStim [24] databases. The pictures were used to generate morphs in 9 steps with Gimp 2.6 (www.gimp.org) and the GAP toolbox, resulting in 11 morphing grades including the original images. Faces were divided into an upper and lower half, with the border defined as being above the nostrils at the bridge of the nose of the middle morph. This border was defined for each face identity and subsequently applied to all of its morphs.

In Experiment 1 there were three visibility conditions (whole face, upper half, lower half). The half of no interest was blurred with a very broad Gaussian filter, preserving only a rough outline of the face, thereby conveying a feeling of wholeness, while rendering the masked features invisible (cf. Fig. 1).

[**Figure 1. Design of Experiment 1.** Illustration of a face morphed from the original fearful (outer left) to the original angry expression (outer right) in 9 intermediary steps, resulting in a total of 11 face morphs; a, whole face; b, upper half intact ('eyes' condition); c, lower half intact ('mouth' condition); due to copyright restrictions, the depicted example is an in-house generated face which was not used in the present experiment.]

2.3 Design

A two-alternatives forced choice identification task was used, in which participants had to decide for each face whether its expression was 'angry' or 'fearful'. Each of the 20 identities was presented in 11 morphing grades. The experiment consisted of two runs with a total of 40 trials per cell, resulting in a resolution of 2.5% for identification performance. The first run was repeated after a short break, to allow for analyses of intra-subject stability. This resulted in a total of 440 trials per masking condition (11 morphs x 20 identities x 2 runs). Pictures were shown with no time limit until a response was given, after which the next stimulus was shown directly, to allow for smooth work-flow. Participants were able to monitor their progress on a status bar presented at the bottom of the screen. Order of stimuli was randomized, the only constraint being that two subsequent trials never contained the same face identity.

Participants had to press the left or right mouse button to indicate whether the target face part showed an angry or fearful expression (button assignment counterbalanced between participants). Experiments were performed using Presentation software (Version 17.1, www.neurobs.com). All experiment files, including raw data, can be found in supplementary files S7 Code and S8 Dataset.

2.4 Analysis

To characterise the participants' performance, a logistic function was fitted to their data. The logistic function ($F_{logistic}(x;\alpha,\beta)=1/[\ 1+\exp(-\beta(x-\alpha))\ ]$) is well-suited to describe observers' performance in psychometric terms [25, p. 82]. It also provides slope information (β parameter) to quantify the steepness of the curve at the point of guessing. An s-shaped logistic function with a steep slope is indicative of high precision in distinguishing between two categories and little uncertainty [25,p. 20]. If categorical perception occurs, the steepest point in the psychometric curve corresponds to the category boundary, as this is the point where two stimuli will be discriminated best [5].

It should be noted that the logistic function might occur for any linear system limited by gaussian noise [26], including decisions based on simple image features such as contrast [27]. However, while the mere presence of a nonlinear or s-shaped curve in an identification task is not sufficient to claim the occurrence of categorical perception, its absence argues strongly against it. Therefore, a linear function ($F_{linear}(x;\alpha,\beta)=\beta*x+\alpha$), as an approximation to a logistic function with uniform slope was fitted to the data. If the logistic and linear function provide a similar fit for a certain condition, this would indicate that observers are unable or at least very poor at distinguishing between two distinct categories. Functions were fitted using an iterative optimization strategy, employing a non-linear least-squares procedure to derive the parameters for intercept and slope which describe the best-fitting function for the raw data. These parameters, together with the sum of squared errors (ss^2) of the best-fitting linear or logistic function were then compared between conditions.

Furthermore, for cross-validation, data from one half of the experiment were used to predict the other half of the data. Since each participant performed every trial twice, data from the first half of the experiment were used to predict the data from the second half (and vice versa). For each condition, one half of the participants' data was used to fit a logistic or linear function. Subsequently, the ss^2 between these data and the left-out half were computed. For example, a logistic function was fit to the first half of the whole face condition and this fitted function was then compared to the raw data of the second half of the experiment, using ss^2.

This allowed assessing whether it is possible to predict what condition a participant was in, given the responses made in a previous run and the function fitted to that data. The more similar two conditions are, the smaller the ss^2 should be. Using sums of squared errors has the advantages of being an intuitive measure of similarity and being virtually assumption-free [28, p. 106].

Data analysis was performed with Python 2.7.9 (www.python.org) using NumPy, SciPy, Pandas, Matplotlib, and the IPython Notebook, all as provided with Anaconda 2.2.0 (Continuum Analytics; docs.continuum.io/anaconda). Full code and output for all analyses can be found in the online supplement. Refer to S1 Code and S2 Code for data import and restructuring; S3 Code for main data analysis; S4 Code for fitting psychometric functions and cross-validation; and to S5 Code for analyses of single face identities. Analyses of Variance (ANOVA) for repeated measure designs were additionally carried out using IBM SPSS 22.

2.5 Results

2.5.1 Comparison of response curves for different masking types

Identification performance was compared across all 11 morphing grades for the three masking conditions (full face, upper face half, lower face half). Participants' performance is illustrated in Fig. 2. A 3x11 repeated-measures ANOVA showed a significant main effect of morphing grade, reflecting that anger responses increase as the faces are morphed towards this expression ($F_{(10,270)}=620.7$, $p<0.001$, $\eta_p^2=0.96$). There was no main effect of masking condition ($F_{(2,54)}=2.3$, $p=0.112$, $\eta_p^2=0.08$), indicating that there is no bias that would shift the responses between conditions. A significant condition-by-morphing grade interaction ($F_{(20,540)}=63.5$, $p<0.001$, $\eta_p^2=0.70$) indicated that the shape of the response curve differs between conditions. Paired t-tests for repeated measures and the Wilcoxon signed rank test as its non-parametric equivalent were carried out to test differences between conditions for all morphing grades. Results were only considered significant if the p-value for both metrics fell below $\alpha=0.05$.

Identification performance was best (i.e. significantly farthest from guessing) for the whole face, compared both to the upper (all *ps*<0.05, except 50% angry morphs) and lower half (all *ps*<0.05, except 60% angry morphs). Furthermore, performance was significantly better for the upper than for the lower half (all *ps*<.05 except for 50% and 60% angry morphs). This indicates that observers were able to make the most accurate decisions when presented with a whole face, followed by face halves containing only the eye region. Performance was worst for only the mouth region.

Judgements of the eye region closely resembled the whole face responses, as indicated by a significantly smaller difference between upper and whole face decisions compared to lower and whole face decisions (all *ps*<0.05 for pairwise comparisons of differences, except for 50% and 60% morphs).

[**Figure 2. Main Results Experiment 1.** Percent angry responses (y-axis) across the 11 morphing grades morphed from fear to anger (x-axis) for the three different visibility conditions (coloured lines); smaller right hand figures depict responses on single-participant level for each condition; 'whole', whole face is visible (no masking); ‘upper', upper face half is visible (lower half is masked); 'lower', lower face half is visible (upper half is masked).]

2.5.2 Cross-validation of conditions

If differences between conditions are genuine and stable, each condition should be able to predict itself best, e.g. the whole face condition of the first half of the experiment should be most similar to the whole face condition of the second half.

This was confirmed by split-half cross-validation with each half of the experiment used once as test and as training set, within each participant. Each condition was significantly closest to its analogue form the other half of the experiment, as expressed in significantly smallest ss^2 for these pairings (all $p<0.01$ for pairwise comparisons).

Furthermore, the upper and whole face conditions were significantly more similar to each other than to the lower half condition ($p<0.01$). This stronger similarity of the eye region to the whole face ratings further corroborates that the upper half carries more information about fear and anger expressions than does the mouth region.

2.5.3 Comparison of the superiority of a logistic fit to the data

A linear and logistic function were both fitted to the data of each participant and ss^2 were computed for all three masking conditions (Fig. 3a), with a better fit indicated by smaller ss^2. The superiority of the logistic over the linear function was significantly greatest for the whole face, followed by the upper face half, and lower face half, respectively ($F_{(2,54)}$=34.1, p<0.001, η_p^2=0.56; all ps<0.001 for pairwise comparisons).

[**Figure 3. Curve Fitting for Experiment 1.** Results of fitting different functions to the data set; red bars indicate 95% confidence intervals of the mean for each condition; line graphs in the background depict the raw data of each participant; black lines indicate that the participant's value for the left-hand condition is numerically bigger than the right-hand value; blue line indicates that the value is smaller for the left-hand compared to right-hand condition; a, superiority of fitting a logistic function to the data, with a higher difference of ss^2 compared to the linear function indicating a better fit; b, slope parameter of a logistic function fitted to the data; higher value indicates steeper slope; c, superiority of a logistic fit when the function is derived from one half of the experiment and ss^2 are computed to the left-out half ; 'whole', whole face is visible (no masking); 'upper', upper face half is visible (lower half is masked); 'lower', lower face half is visible (upper half is masked).]

2.5.4 Comparison of the slopes of the logistic function between conditions

When fitting a logistic function to each participant's data for each condition, the steepness of the slopes can be compared as measure for the presence of a category boundary, with a steep slope indicating stronger categorical perception (Fig. 3b).

Results show that the slopes were steepest for the whole face condition, followed by the upper and lower half ($F_{(2,54)}$=128.2, p<0.001, η_p^2=0.83; all ps<0.001 for pairwise comparisons), indicating a more abrupt shift from fear to anger responses when the whole face is visible and a smooth shift in responses when viewing the lower face half.

2.5.5 Cross-validation of response curves for different masking types

Left-out data from one half of the experiment were predicted using data from the other half of the experiment with the two runs of the experiment used for cross-validation. Each condition was modelled for one half of the experiment using a linear or logistic function fitted to these data. These training sets were then compared to the left-out half of the experiment to check how the models compare to each other in predicting new data.

The results of the cross-validation analysis (Fig. 3c) show that the superiority of a logistic over a linear function for predicting left-out data is largest for the whole face condition and weakest when only the lower face half is visible ($F_{(2,54)}$=21.0, p<0.001, $\eta_p^2=0.44$; ps<0.05 for all pairwise comparisons).

3. EXPERIMENT 2

3.1 Methods

Mean age of participants was 24 years (range 19-29). Of the 30 participants, 24 were female. Participants received course credit or 7EUR for participation and gave oral informed consent before starting the experiment.

The same boundary to divide faces into lower and upper halves as in experiment 1 was used and all other design choices were identical to experiment 1, with the following exceptions: This time 100% fearful or 100% angry expression halves were combined with the morphed pictures to create the composite face illusion. There were four conditions: 'lower half 100% angry' (Fig. 4a), 'lower half 100% fearful (Fig. 4b), 'upper half 100% angry' (Fig. 4c) and 'upper half 100% fearful' (Fig. 4d). The complementary half was presented in 11 morphing steps, and was the part which the observers were asked to identify as either angry of fearful. This part was always framed by a red square (cf. Fig. 4). Participants were explicitly instructed to focus only on the framed face half and to ignore the other half. Unlike experiment 1, order of conditions was blocked, to make it easier for the participants to focus on the part of interest. Each block required either always judging the upper half or always judging the lower half, and consisted of 440 trials, with block order counterbalanced between participants. Within blocks, order of the 20 face identities, the 11 morphed face halves and whether the to-be-ignored half was fearful or angry were randomised.

[Figure 4. Design of Experiment 2 (Composite Face). Illustration of a face morphed from the original fearful (outer left) to the original angry expression (outer right) in 9 intermediary steps, resulting in a total of 11 face morphs; a, eye judgements with 100% angry lower face half; b, eye judgements with 100% fearful lower half; c, mouth judgements with 100% angry upper face half; d, mouth judgements with 100% fearful upper half; conditions a and b or conditions c and d were always presented in one block, to aid participants in focusing on one face half only; due to copyright restrictions, the depicted example is an in-house generated face which was not used in the present experiment.]

One participant reported to have only focused on the mouth region throughout the experiment, ignoring which face half was currently framed. These data were excluded, leaving 29 data sets for analysis.

3.2 Results

3.2.1 Comparison of response curves for different composite faces

Depending on the viewing condition, responses for the attended face half were biased in the direction of the to-be-ignored face half (Fig. 5a).

A 2x2x11 ANOVA with the factors upper/lower face half, anger/fear distractor and morphing grade (fear to anger) revealed a main effect of morphing grade ($F_{(10,280)}$=213.7, p<0.001, η_p^2=0.88), a main effect of emotion ($F_{(1,28)}$=109.6, p<0.001, η_p^2=0.80) and significant interactions between all factors (all *ps*<0.001). The main effect of emotion in the distractor half reflects that responses for the eye region were shifted towards anger when the lower face half was angry, and towards fear when the lower half was fearful (p<0.01 for all 11 morphing grades). The same bias occurred for the judgements of the mouth region, when the upper half was either angry or fearful (Fig. 5c; all p<0.001). The interaction effect between emotion and face half ($F_{(1,28)}$=26.0, p<0.001, η_p^2=0.48) reflects that the bias was significantly stronger for mouth judgements than for eye judgements, indicating that the upper face half exerts a stronger influence on the lower half than vice versa. Across all 11 morphing grades, fearful and angry distracting halves led to bigger differences in mouth judgements, compared to the differences in eye judgements (all $p<0.001$, for pairwise comparisons of differences).

[Figure 5. Main Results for Composite Faces. Percent angry responses (y-axis) across the 11 morphing grades morphed from fear to anger (x-axis) for the four different composite face conditions (coloured lines); left-hand figures depict mean results with 95% confidence intervals on group level; right-hand figures depict responses on single-participant level for each condition; 'lower half fearful', observers judge the upper half, while the lower half is always 100% fearful; 'lower half angry', observers judge the upper half, while the lower half is always 100% angry; 'upper half fearful', observers judge the lower half, while the upper half is always 100% fearful; 'upper half angry', observers judge the lower half, while the upper half is always 100% angry.]

3.2.2 Cross-validation of conditions

Split-half cross-validation with each half of the experiment used once as test and as training set showed that each condition was significantly closest to its analogue form the other half of the experiment (all $p<0.01$ for pairwise comparisons), indicative of reliable differences between conditions.

3.2.3 Comparison of the superiority of a logistic fit to the data

The superiority of a logistic function over a linear function was significantly stronger for eye judgements as compared with mouth judgements (Fig 6a; main effect of face half $F_{(1,28)}=13.7$, $p<0.001$, $\eta_p^2=0.33$; all $p<0.05$ for pairwise comparisons), indicating that when observers judge the upper face half, their response function is significantly more s-shaped that when they judge the lower face half.

[**Figure 6. Curve Fitting for composite faces.** Results of fitting different functions to the data set; red bars indicate 95% confidence intervals of the mean for each condition; line graphs in the background depict the raw data of each participant; 'rating upper/lower fearful', observers judge the upper half, while the lower half is always 100% fearful; 'rating upper/lower angry', observers judge the upper half, while the lower half is always 100% angry; 'rating lower/upper fearful', observers judge the lower half, while the upper half is always 100% fearful; 'rating lower/upper angry', observers judge the lower half, while the upper half is always 100% angry; for more information please refer to the legend of figure 3.]

3.2.4 Comparison of the slopes of the logistic function between conditions

When fitting a logistic function to each participant's data for each of the three conditions, the steepness of the slopes was significantly higher for upper than for lower face half judgements (Fig. 6b; main effect of face half; $F_{(1,28)}$=81.4, p<0.001, $\eta_p{}^2$=0.74; all ps<0.001 for pairwise comparisons), indicative of a more abrupt change in fear/anger decisions when viewing the upper half. However, since lower half judgements were biased by the eye region to the point where for many participants performance never crossed 50% guessing, slope parameters for the mouth regions are difficult to interpret.

3.2.5 Cross-validation of response curves for different masking types

Analogous to experiment 1, results of the cross-validation analysis show that the superiority of a logistic over a linear function for predicting left-out data is larger for upper half versus lower half judgements (Fig. 6c), but only significant when the ignored face half was fearful (p<0.05). The difference was only numerically superior when an anger mouth was present (significant interaction effect $F_{(1,28)}$=4.8, p=0.038, $\eta_p{}^2$=0.15).

4. RESULTS FOR SINGLE FACE IDENTITIES

Since 20 different face identities were used as stimuli in the experiments, the present results may underestimate or blur the sigmoid shape of the response functions, due to averaging over a heterogeneous set of stimuli with shifted intercepts.

Therefore, data of both experiments were re-analysed by fitting a logistic function to the responses to each of the 20 individual face identities, averaged across participants. Subsequently, the intercepts were shifted to be uniformly at 50%, so that each curve is at 50% guessing (y-axis) when the face is 50% angry (x-axis). The necessary shift was determined from the fitted logistic functions and then applied to the underlying raw data, allowing to compare how the curves would be shaped, given equal intercepts for all face identities. The results are illustrated in Fig. 7 and are described in-depth in S5 Code. Visual inspection of the data suggests that a sigmoid function clearly emerges only for whole faces and upper halves in experiment 1 and for upper halves in experiment 2 when accounting for potential variability between faces.

[Figure 7. correction for different intercepts between face identities. Main analyses for both experiments after modelling responses to each face identity individually and shifting the intercepts to be uniformly at 50%; a, results for experiment 1 (masking); b, results for experiment 2 (composite faces); note that a different number of data points is available for each bin on the x-axis, and therefore estimates at the extremes are less reliable; especially in experiment 2, mouth judgements were strongly biased, with few responses at 50% guessing; please refer to Supplement S5 for more details on the analysis procedure.]

5. DISCUSSION

The present study set out to investigate what part of the face is most diagnostic for categorical perception of fear-anger morphed expressions, and how this information is integrated when presented in a composite face.

Experiment 1 showed that the upper face half is more diagnostic than the lower half for deciding between fearful and angry expressions in an identification task. The upper half alone was sufficient for high identification performance, closely reflecting the whole face condition. Its perception could also be well-described by a logistic function with a steep category boundary, providing converging evidence for categorical perception based on the upper face half alone. When viewing the lower half, responses were significantly closer to chance, were less s-shaped and had a comparably shallower slope.

In experiment 2, it was expected that the to-be-ignored halves of the face would bias performance, and that regions that have proven most diagnostic in experiment 1 should exert a stronger biasing influence over complementary face parts than vice versa. Accordingly, results indicated that both eye and mouth regions bias judgements on the respective other face half, with strongest influences exerted by the eyes, leading to what might be considered a breakdown of performance for mouth judgements. This effect is noteworthy, as information in the mouth region is equally present in both experiments. Nevertheless, when instructed to evaluate the mouth and ignore the eyes in a composite face, the participants' ratings reflect the expression of the eyes numerically more strongly than they reflect mouth judgements. This means that a cue that is reasonably useful for expression recognition in isolation, will be of almost no use when exposed to the influence of a strongly diagnostic cue.

The influence of the mouth on eye judgements should be considered equally striking, since it illustrates that even a very diagnostic cue that allows for recognition levels almost as high as a whole face, can be subject to significant bias form a substantially weaker cue. This suggests that even when analytical processing of a single feature is sufficient to allow for categorical perception, human observers seem to integrate features into a whole whenever possible.

The dominance of the eye region in the present experiments is well in line with masking studies which point towards its prominent role for both fear and anger recognition [13]. Studies with neurological patients also indicate a specific importance of the eyes in fear detection [22]. Also, when asked to evaluate expressions of emotion, observers most readily fixate the eyes [14], irrespective of the emotion in question.

Regarding identity recognition, there are many studies that demonstrate the importance of the eyes. For example, when a face has to be learned in a training run and later distinguished from distractors differing in one feature, it is easier to recognize identity changes due to differences in the eyes, compared to nose or mouth changes [29]. Masking out the eyes may even represent a legal requirement in some contexts for ensuring anonymity. Although the scientific basis for the effectiveness of this procedure is equivocal at best [cf. 30], it illustrates the paramount importance we intuitively place on the eye region. Regarding mental states, it is also commonly assumed that they can be assessed best from the eye region, as reflected for example in the 'reading the mind in the eyes' test [31]. Accordingly, one must ask whether the eye advantage reported in the present study can be generalised to other features and other expressions of emotion.

Even though the eyes may be most likely fixated when judging emotions from intact faces [14], this does not necessarily imply that most information is coded in the eye region. This becomes obvious when considering the studies carried out with patient SM, who exhibits major difficulties at recognizing fear from faces, due to a failure to fixate the eyes [22]. Nevertheless, she is virtually unimpaired at recognizing all other basic emotions. This indicates that it must be possible to recognize those emotions utilising other diagnostic features. Accordingly, masking studies [13] and coding systems for basic emotion expressions [10], suggest that each expression can be best recognized from a set of features specific to that emotion and that focusing on the eyes cannot be the optimal strategy in all cases.

Therefore, regarding the results of the present experiments, we would predict that the basic mechanisms outlined here will be replicated with other facial expressions, but perhaps not for the same features. For example, happy-surprised pairings should be best recognized from the mouth, as both involve mainly muscles of the lower face half [10,13,32]. Interestingly, eye-tracking studies found that more fixations on the eye regions correlates with improved emotion recognition for all expressions [14,33] and argue that fixating the eyes reflects a holistic processing strategy. One could test this assumption using happy-surprise pairings and the composite face task from experiment 2. If holistic processing is associated with fixating the eyes, then eye judgements should be easier in a composite face task, even if a strongly diagnostic distracting lower face half is present.

The present findings should not only be extended to investigate other expressions, but also further elaborated for the currently used fear-anger pairing, to shed light on the basic perceptual mechanisms at work. The employed identification task could be complemented by a discrimination task. There, it should be easier to discriminate two images when they are on opposing sides of the category boundary, compared with being both sampled from within the same category [5,6]. Thus, the sigmoid curves from the present experiments could be used to predict the discrimination performance in a follow-up experiment. As differences in slope between conditions can reflect differences in variance of the underlying noise distribution or the uncertainty of the observer [34,35], it would be important to establish that a between-category advantage [5] is indeed present at the point where the psychometric function is steepest in the identification task.

On the basis of the present data from experiment 1, we would therefore predict that discrimination performance at the category boundary should be higher for eye than mouth judgements.

In the composite face task, the peak of the discrimination function should be shifted to the right and left, respectively, when judging the eyes, in accordance with the intercept shifts due to the mouth expressions. Finally the discrimination function should be almost flat when judging the mouth in a composite face, as there is virtually no crossing of a category boundary in the present data.

To further generalise the present findings regarding fear-anger pairings, averaged faces could be used, which have the advantage that idiosyncrasies would average out while expressions prototypical for a certain emotion would be emphasized [36, p. 76]. While studies on categorical perception are often performed with a very limited number of identities (<=4; [cf. 4,7,8,21]), the 20 different faces used in the present experiments should be considered a valuable step towards greater generalisability. However, the fact that the mouth provided so little diagnostic information might be partly due to the fact that 12 out of the 20 face identities expressed the emotions with a closed mouth. While this might have reduced variance in the expressiveness of the lower face half, both the pressing of the lips as well as the baring of the teeth are valid anger expressions [10,12]. Of the two, lip pressing is probably much more frequent in everyday life. Also, when making statements about the importance of either eyes or mouth for recognizing an expression, it should be kept in mind that these features were rather crudely operationalised as the upper and lower face half, respectively. While such a division is standard practice in the field [18,21,37], the muscles around the nose have likely contributed to the judgements of both face halves, and the eyebrows [38] and forehead [11] might have contributed to judgements of the upper face half. While the sclera of the eyes seems crucial for fear detection [39], anger judgements might depend more on the surrounding muscles of the eyes and the eyebrows [13,38], especially since patient SM shows no difficulty at recognizing anger [22].

Another interesting question is to what degree the failure to concentrate on one half of a composite face can be voluntarily overridden, especially since previous work has shown that emotion recognition improvements are associated with increased fixation of the eyes [33]. This could be investigated by introducing training blocks before the experiment proper, a forced minimum exposure time to discourage spontaneous responses, biofeedback from an eye-tracking device helping to fixate on a specific face part, and rewards for correct responses.

To summarise, the present studies show that identification of fear and anger in morphed faces relies heavily on the eye region, corroborating previous research [10,13,14,19]. Expanding previous work, the study shows that observers can perform categorical perception even when viewing only a single face part. For the employed fear-anger pairings, the eye region was sufficient to allow for decisions with a high degree of certainty. This demonstrates that categorical perception can emerge based on single features only, adding to our understanding of how categorization works in complex stimuli such as faces.

Even though the eye region was almost as diagnostic as the whole face, it was not immune to biasing influences in a composite face task. When a full face with conflicting expressions was shown, the psychometric curves for eye judgements were shifted in accordance with the expression in the lower half of the face. For the first time, we can therefore show that holistic perception of a full face can be explained based on the diagnostic value of its constituent parts. In addition, given that observers were aware that a holistic processing strategy impaired their performance, this illustrates that humans involuntarily process faces as a gestalt when trying to make categorical decisions about its expression.

ACKNOWLEDGEMENTS

We would like to thank Maleen Fiddicke for help with data collection.

REFERENCES

1. Keltner D, Haidt J. Social Functions of Emotions at Four Levels of Analysis. Cogn Emot. 1999;13: 505–521. doi:10.1080/026999399379168

2. Ekman P. Are there basic emotions? Psychol Rev. 1992;99: 550–553. doi:10.1037/0033-295X.99.3.550

3. Du S, Tao Y, Martinez AM. Compound facial expressions of emotion. Proc Natl Acad Sci. 2014;111: E1454–E1462. doi:10.1073/pnas.1322355111

4. Young AW, Rowland D, Calder AJ, Etcoff NL, Seth A, Perrett DI. Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. Cognition. 1997;63: 271–313. doi:10.1016/S0010-0277(97)00003-6

5. Fugate JMB. Categorical Perception for Emotional Faces. Emot Rev. 2013;5: 84–89. doi:10.1177/1754073912451350

6. Harnad SR. Categorical Perception. Categorical perception: the groundwork of cognition. Cambridge; New York: Cambridge University Press; 1990.

7. Calder AJ, Young AW, Perrett DI, Etcoff NL, Rowland D. Categorical Perception of Morphed Facial Expressions. Vis Cogn. 1996;3: 81–118. doi:10.1080/713756735

8. De Gelder B, Teunisse J-P, Benson PJ. Categorical Perception of Facial Expressions: Categories and their Internal Structure. Cogn Emot. 1997;11: 1–23. doi:10.1080/026999397380005

9. Etcoff NL, Magee JJ. Categorical perception of facial expressions. Cognition. 1992;44: 227–240. doi:10.1016/0010-0277(92)90002-Y

10. Ekman P, Friesen WV, Hager JC. Facial action coding system. Salt Lake City UT Hum Face. 2002;

11. Ekman P, Friesen WV. Unmasking the face: a guide to recognizing emotions from facial clues. Cambridge, MA: Malor Books; 1975.

12. Matsumoto D, Ekman P. Facial expression analysis. Scholarpedia. 2008;3: 4237. doi:10.4249/scholarpedia.4237

13. Smith ML, Cottrell GW, Gosselin F, Schyns PG. Transmitting and Decoding Facial Expressions. Psychol Sci. 2005;16: 184–189. doi:10.1111/j.0956-7976.2005.00801.x

14. Guo K. Holistic Gaze Strategy to Categorize Facial Expression of Varying Intensities. Verdejo García A, editor. PLoS ONE. 2012;7: e42585. doi:10.1371/journal.pone.0042585

15. Rossion B. Picture-plane inversion leads to qualitative changes of face perception. Acta Psychol (Amst). 2008;128: 274–289. doi:10.1016/j.actpsy.2008.02.003

16. Sergent J. An investigation into component and configural processes underlying face perception. Br J Psychol. 1984;75: 221–242. doi:10.1111/j.2044-8295.1984.tb01895.x

17. McKone E, Yovel G. A single holistic representation of spacing and feature shape in faces. J Vis. 2010;8: 163–163. doi:10.1167/8.6.163

18. Young AW, Hellawell D, Hay DC. Configurational information in face perception. Perception. 1987;16: 747–759. doi:10.1068/p160747

19. Tanaka JW, Kaiser MD, Butler S, Le Grand R. Mixed emotions: Holistic and analytic perception of facial expressions. Cogn Emot. 2012;26: 961–977. doi:10.1080/02699931.2011.630933

20. Calder AJ, Jansen J. Configural coding of facial expressions: The impact of inversion and photographic negative. Vis Cogn. 2005;12: 495–518. doi:10.1080/13506280444000418

21. Chen M-Y, Chen C-C. The contribution of the upper and lower face in happy and sad facial expression classification. Vision Res. 2010;50: 1814–1823. doi:10.1016/j.visres.2010.06.002

22. Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. Nature. 2005;433: 68–72. doi:10.1038/nature03086

23. Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces (KDEF). CD ROM Dep Clin Neurosci Psychol Sect Karolinska Institutet. 1998; 91–630.

24. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: Judgments from untrained research participants. Psychiatry Res. 2009;168: 242–249. doi:10.1016/j.psychres.2008.05.006

25. Kingdom FAA, Prins N. Psychophysics: a practical introduction. London: Academic; 2010.

26. Duda RO, Hart PE, Stork DG. Pattern classification [Internet]. 2001. Available: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=531488

27. Foley JM, Legge GE. Contrast detection and near-threshold discrimination in human vision. Vision Res. 1981;21: 1041–1053. doi:10.1016/0042-6989(81)90009-2

28. Hilborn R. The ecological detective: confronting models with data. Princeton, NJ: Princeton University Press; 1997.

29. Tanaka JW, Farah MJ. Parts and wholes in face recognition. Q J Exp Psychol Sect A. 1993;46: 225–245. doi:10.1080/14640749308401045

30. Newton EM, Sweeney L, Malin B. Preserving privacy by de-identifying face images. IEEE Trans Knowl Data Eng. 2005;17: 232–243. doi:10.1109/TKDE.2005.32

31. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. J Child Psychol Psychiatry. 2001;42: 241–251. doi:10.1111/1469-7610.00715

32. Leppänen JM, Hietanen JK. Is there more in a happy face than just a big smile? Vis Cogn. 2007;15: 468–490. doi:10.1080/13506280600765333

33. Pollux PMJ, Hall S, Guo K. Facial Expression Training Optimises Viewing Strategy in Children and Adults. Barton JJS, editor. PLoS ONE. 2014;9: e105418. doi:10.1371/journal.pone.0105418

34. Ashby FG, Townsend JT. Varieties of perceptual independence. Psychol Rev. 1986;93: 154–179.

35. Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. J Opt Soc Am A. 1985;2: 1508–1532.

36. Bruce V, Young AW. Face perception. London ; New York: Psychology Press; 2012.

37. Grand RL, Mondloch CJ, Maurer D, Brent HP. Impairment in Holistic Face Processing Following Early Visual Deprivation. Psychol Sci. 2004;15: 762–768. doi:10.1111/j.0956-7976.2004.00753.x

38. Tipples J, Atkinson AP, Young AW. The eyebrow frown: A salient social signal. Emotion. 2002;2: 288–296. doi:10.1037/1528-3542.2.3.288

39. Whalen PJ. Human Amygdala Responsivity to Masked Fearful Eye Whites. Science. 2004;306: 2061–2061. doi:10.1126/science.1103617

SUPPLEMENT

S1 Code. Data import for experiment 1

S2 Code. Data import for experiment 2

S3 Code. Main analysis and plotting

S4 Code. Curve fitting and cross-validation

S5 Code. Additional analyses for face identities

S6 Code. Complete analysis code from S1-S5 in executable IPython Notebook format

S7 Code. Complete code for experiment presentation in Neurobs Presentation format

S8 Dataset. Complete raw data of all participants

Reviewer #1:

As the author said in Discussion, their result was obtained with angry and fearful images, it may not be able to generalize to other expressions, such as happy, surprised or sad. This limitation should be reflected in the title and abstract.

Thank you for your suggestion. We have changed the title of our manuscript to

> „Categorical Perception of Fear and Anger Expressions in Whole, Masked and Composite Faces.“.

Also, there are two passages in the abstract which make clear that our study deals with fearful and angry faces only:

> „We report results from two experiments with morphed **fear-anger expressions** [...]“
>
> abstract, 3rd paragraph

> „The present study shows that identification of **fear and anger** in morphed faces relies heavily on information from the upper half [...]“
>
> abstract, last paragraph

In addition, we have added the following clarification to the abstract:

> „Important questions are, whether there are single diagnostic regions in the face that drive categorical perception for certain parings of emotion expressions [...]”
>
> abstract, 2nd paragraph

The main argument of this paper, as stated in Section 2.4, is based on the assumption: “The linear function (Flinear(x;α,β)=β*x+α) should correspond best to changes in low-level features and would be a good fit if observers' performance closely followed the basic image properties.

The logistic function (Flogistic(x;α,β)=1/[1+exp(-β(x-α))]) is well suited to describe observers' performance in psychometric terms [25, p. 82] and should provide an optimal fit if the observers engage in categorical perception". The authors made this assumption without an explanation or a reference (the cited Kingdom & Prins book was about logistic function, not its implications). Actually, this assumption is wrong. The logistic psychometrics function occurs for any linear system limited by Gaussian noise (see Green & Swets, 1966, Signal Detection Theory; Duda & Hart, 1973, Pattern Classification and Scene Analysis, and the Kingdom & Prins book cited by the authors), even the one whose decision is based on simple image features, such as contrast (e.g., Foley & Legge, 1981, Vision Res., 21, 1041-53). Hence, whether the psychometric function is linear or logistics cannot help the authors to decide whether the observer is based on "low-level features" or "categorical perception" (By the way, neither term were defined in the paper).

We agree and would like to thank the reviewer for pointing out this important issue. There certainly is no one-to-one mapping between categorical perception and an s-shaped response function. However, we believe that the presence of an s-shaped logistic function is a necessary condition to claim categorical perception. For example, the assumption of categorical perception for syllables (e.g. ba/pa) or colours would be incompatible with a strictly linear increase in the response function. Although the logistic function can arise due to different reasons, as pointed out by the reviewer, its absence would provide strong evidence against categorical perception (i.e. in the sense of the modus tollens).

Accordingly, we wrote in our manuscript:

> *"[...] the linear changes in low-level features are insufficient to describe the changes in perception, which instead follow an s-shaped function"*
>
> and

“The linear function […] would be a good fit if observers' performance closely followed the basic image properties […] The logistic function [...] should provide an optimal fit if the observers engage in categorical perception”

We believe that these statements are logically sound, as they do not claim that we want to dichotomize between categorical perception and a strict following of low-level features. There was a statement in the results section which we agree was problematic:

“Results show that the slopes were steepest for the whole face condition […] Thus, categorical perception was strongest when all face features were visible [...]”

This too far-reaching claim has been removed from the manuscript.

Also, we have added a discussion of these issues in the methods section of our manuscript:

“It should be noted that the logistic function might occur for any linear system limited by gaussian noise [26], including decisions based on simple image features such as contrast [27]. However, while the mere presence of a nonlinear or s-shaped curve in an identification task is not sufficient to claim the occurrence of categorical perception, its absence argues strongly against it.”

(2.4 Analysis, 2nd paragraph)

We have also added explicit definitions of the term “low-level features” to the introduction of our manuscript (in the respective paragraph, a definition of “categorical perception” was already given):

“Such categorical perception is demonstrated by identification tasks where linear changes in low-level features (e.g. basic visual properties like colour or shape) lead to non-linear changes in perception, which are best described by a sigmoid function with a steep category boundary. In general, categorical perception refers to the phenomenon that objects from the same category will be perceived as more similar than objects from different categories [5], despite being equally far apart from each other on a given physical dimension.”

(Introduction, 3rd paragraph)

The second reason why a comparison between linear and logistic fits is meaningless is that any linear psychometric function can be considered as an approximation to a logistic function with limited range.

We agree that our initial analyses with direct comparisons between the fit of a logistic and linear function should be omitted. As it still is of importance to us to provide objective metrics for how s-shaped the logistic function is for each condition, we now adopt the rationale of comparing the degree of superiority a logistic fit between conditions. This avoids attaching meaning to the fact that the logistic function per se provides a superior fit over the linear one. The metric of interest instead is, for which condition this superiority is most pronounced. We feel that this is an important reframing of our analysis rationale and hope that the reviewer agrees.

This amended analysis rationale is made explicit in the methods section:

> *"[…]* while the mere presence of a nonlinear or s-shaped curve in an identification task is not sufficient to claim the occurrence of categorical perception, its absence argues strongly against it. *Therefore, a linear function ($F_{linear}(x;\alpha,\beta)=\beta*x+\alpha$), as an approximation to a logistic function with uniform slope was fitted to the data. If the logistic and linear function provide a similar fit for a certain condition, this would indicate that observers are unable or at least very poor at distinguishing between two distinct categories."*

(2.4 Analysis, 2nd paragraph)

The results section has been amended, with the respective sections only reporting the degree of superiority of the logistic function between conditions. Figures 3 and 6 have also been modified accordingly (cf. sections 3.2.3 -3.2.5 for experiment 1 and sections 3.2.3 -3.2.5 for experiment 2) .

The authors also made a considerable effort describing the slope of the psychometric function.

Their rationale was described in Section 4.2 "...slope information (β parameter), to quantify how steep the category boundary is for a certain condition." There are many reasons underlying a specific slope of a psychometric function. A shallow slope may be caused by a large variance of the underlying noise distribution (Ashby & Townsend, 1986, Psych. Rev., 93, 154-79) or a low level of uncertainty in the system (Pelli, 1985, JOSAA, 2, 1508-32; Tyler & Chen, 2000, Vision Res., 40, 3121-44). It is not necessary reflecting the steepness of the boundary.

We agree that alternative explanations should always be considered. Therefore, we added the reviewer's concerns to the discussion of our manuscript:

> *"The employed identification task could be complemented by a discrimination task. There, it should be easier to discriminate two images when they are on opposing sides of the category boundary, compared with being both sampled from within the same category [5,6]. Thus, the sigmoid curves from the present experiments could be used to predict the discrimination performance in a follow-up experiment. As differences in slope between conditions can reflect differences in variance of the underlying noise distribution or the uncertainty of the observer [34,35], it would be important to establish that a between-category advantage [5] is indeed present at the point where the psychometric function is steepest in the identification task."*
>
> *(Discussion section, 9th paragraph)*

However, there are two main reasons why we are confident that our use of slope parameters is sensible:

Firstly, there is a compelling relationship between the steepness of the slope in Exp. 1 and the performance in the composite face task in Exp. 2. Namely, the steeper the slope in Exp. 1, the more the respective face part is resilient to biasing influence. Likewise, a shallow slope in Exp. 1 leads to a strong susceptibility to distractors in Exp. 2. Therefore, the steepness of the slope indicates how robustly observers can distinguish between two categories in the presence of distractors.

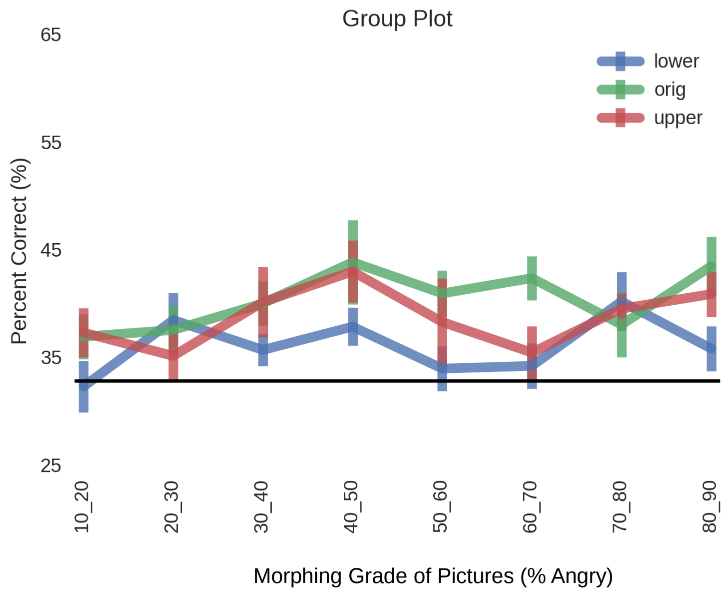
Secondly, there is a substantive body of literature linking the steepness of the slope to performance in a discrimination task, e.g. an AB-X or same-different task (Etcoff & Magee 1992 Cognition; Calder et al. 1996 Visual Cognition, Young et al. 1997 Cognition). In the discussion section of our manuscript, we addressed the need to validate the current data with a discrimination task:

> *"On the basis of the present data from experiment 1, we would therefore predict that discrimination performance at the category boundary should be higher for eye than mouth judgements."*

(Discussion section, 9th paragraph)

Since our submission of the current manuscript, we have acquired pilot data from a three-alternatives forced-choice discrimination task. In this task, three faces are shown with one oddball face differing in its expression by 10%. While we are still in the process of designing the task and the available data are only tentative, they fit our predictions. They illustrate that with full face and upper face halves observers show best performance for the middle morphs, where the psychometric curve is steepest, while discrimination performance is uniformly low for the lower face half (cf. Figure R1 below), where the slope is shallow. The discrimination performance seems to closely mimic the steepness of the curves in the current identification task.

Figure R1. Pilot Data From a Discrimination Task



A three-alternatives forced choice task was employed, where three faces are shown simultaneously and one face is an oddball which has to be identified. The oddball always differed in its morphing grade by 10%. Pilot data of N=9 participants (with design modifications and unequal trial numbers between participants);
“orig”= full face; “lower” = only lower face half is visible; “upper”= only upper face half is visible.

We feel that these results help to support our claim that the steepness of the slope in an identification task carries valuable information about the ability of an observer to engage in categorical perception.

Statistical procedures for model comparison and cross validation were not described. While the authors provided Python codes in the Supplementary material, it is not enough. Not everyone uses Python. The authors should describe their methods in the text such that the readers can get an idea about the method they used without consulting supplementary material.

We apologize if the methods section was not comprehensive enough. It was not our intention that readers would have to study the supplement to properly understand the statistical procedures. We have now added the following paragraphs to the methods section:

Firstly, to better explain the use of sums of squared errors for model comparison:

> *"Functions were fitted using an iterative optimization strategy, employing a non-linear least-squares procedure to derive the parameters for intercept and slope which describe the best-fitting function for the raw data. These parameters, together with the sum of squared errors (ss^2) of the best-fitting linear or logistic function were then compared between conditions."*

(2.4 Analysis, 3rd paragraph)

Secondly, to explain the cross-validation the following paragraph was extended:

> *"Furthermore, for cross-validation, data from one half of the experiment were used to predict the other half of the data. Since each participant performed every trial twice, data from the first half of the experiment were used to predict the data from the second half (and vice versa). For each condition, one half of the participants' data was used to fit a logistic or linear function. Subsequently, the ss^2 between these data and the left-out half were computed. For example, a logistic function was fit to the first half of the whole face condition and this fitted function was then compared to the raw data of the second half of the experiment, using ss^2."*

(2.4 Analysis, 4th paragraph)

Reviewer #2: This manuscript addresses a topical research question and provides interesting support for existing knowledge about the importance of both analytic and holistic processing of faces for facial expression categorization.

General comment: As acknowledged by the authors, evidence for categorical perception of facial expressions is numerous. The authors should highlight more explicitly what their

findings add to existing knowledge and theories. In the introduction, theory-based predictions could be made.

We have made our goals and predictions more explicit in the introduction:

> *"Previous masking studies with full-blown fear and anger expressions have shown that both expressions are mainly recognized from the eye region [13] and studies with neurological patients also point to a special role of the eyes for recognizing fear [22]. Therefore, one can hypothesize that categorical perception of fear and anger will rely mostly on the features in the upper half of the face. In line with research on the composite face illusion [19,20], the more informative half of the face should also dominate the perception of a full face."*
>
> (Introduction, paragraph before last)

We have also completely re-worked the last paragraphs of the discussion section to describe the value of the findings for our theoretical understanding of face perception:

> *"Expanding previous work, the study shows that observers can perform categorical perception even when viewing only a single face part. For the employed fear-anger pairings, the eye region was sufficient to allow for decisions with a high degree of certainty. This demonstrates that categorical perception can emerge based on single features only, adding to our understanding of how categorization works in complex stimuli such as faces. Even though the eye region was almost as diagnostic as the whole face, it was not immune to biasing influences in a composite face task. When a full face with conflicting expressions was shown, the psychometric curves for eye judgements were shifted in accordance with the expression in the lower half of the face. For the first time, we can therefore show that holistic perception of a full face can be explained based on the diagnostic value of its constituent parts. In addition, given that observers were aware that a holistic processing strategy impaired their performance, this illustrates that humans involuntarily process faces*

as a gestalt when trying to make categorical decisions about its expression."

(Discussion Section, last paragraph)

Statistical analysis: The use of a large number of t-tests without correction for inflated Type 2 **error is not appropriate. The data in most sections can be analysed using ANOVA (excluding the curve-fit analyses at the beginning of sections 2.5.3 and 3.2.3) and post-hoc corrections should be used for analysis of significant main or interaction effects. An indication of effect size of effects should also be included (eta squared).**

Thank you for this suggestion. We now provide ANOVA analyses and hope that readers will benefit from them. However, we would like to point out that all information provided by an ANOVA was already conveyed in the original analyses. Each section of the results now features analyses of a repeated-measures ANOVA as well as the original pairwise tests for more focused and directional testing.

Changes have been made throughout the results section and are marked in red color. One example of the use of repeated-measures ANOVA is given here:

> *"Identification performance was compared across all 11 morphing grades for the three masking conditions (full face, upper face half, lower face half). Participants' performance is illustrated in Fig. 2. A 3x11 repeated-measures ANOVA showed a significant main effect of morphing grade, reflecting that anger responses increase as the faces are morphed towards this expression ($F_{(10,270)}$=620.7, p<0.001, η_p^2=0.96). There was no main effect of masking condition ($F_{(2,54)}$=2.3, p=0.112, η_p^2=0.08), indicating that there is no bias that would shift the responses between conditions. A significant condition-by-morphing grade interaction ($F_{(20,540)}$=63.5, p<0.001, η_p^2=0.70) indicated that the shape of the response curve differs between conditions."*

(Section 2.5.1, first paragraph)

We have performed a large number of paired t-tests, due to the fact that we have a very fine-grained parametric modulation of our independent variable (11 morphing steps). We believe that it is of interest to provide the readers with precise information for which morphing grades the masking conditions differ. Therefore, we think that the t-test do add value to the manuscript and help the reader in interpreting the figures.

There are two kinds of corrections we could do. A correction for Type 1 error (e.g. a Bonferroni-correction), would basically penalise any design with fine-grained modulation of an independent variable. As it is not the case that the statistical evidence for an effect weakens as more and more significant differences are found, we find this inappropriate. If two conditions differ in 11 out of 11 levels of a variable (in a systematic and plausible pattern), this alone must be considered a highly significant effect under a binomial distribution (e.g. like a coin coming up heads 11 times in a row).

A correction for Type 2 error would mean that we should increase our alpha to be more sensitive to smaller differences (e.g. by using a p-value higher than 0.05). We believe that the best we can do to avoid Type 2 error is to have a good signal-to-noise ratio and collect a large sample. We feel that this is the case in both of our experiments, given that (a) the effects are robust and visible on a single subject level and (b) the effects can be replicated when splitting the data in half.

Therefore, while we share the concerns about both Type-1 and Type-2 errors with the reviewer, we are convinced that our design, our comprehensive analyses and the results speak in favour of robust and reproducible effects.

There are a few sentences that seem out of context or seem non-informative (e.g. Discussion,

second page, par 4: "Anonymization of pictures…").

We would like to thank the reviewer for pointing this out. We have re-phrased the referenced part to make our point clearer and establish a better connection to the previous paragraph:

> "*Regarding identity recognition, there are many studies that demonstrate the importance of the eyes. For example, when a face has to be learned in a training run and later distinguished from distractors differing in one feature, it is easier to recognize identity changes due to differences in the eyes, compared to nose or mouth changes [29]. Masking out the eyes may even represent a legal requirement in some contexts for ensuring anonymity. Although the scientific basis for the effectiveness of this procedure is equivocal at best [cf. 30], it illustrates the paramount importance we intuitively place on the eye region.* "
>
> (Discussion Section, 6th paragraph)

Discussion: Similar to the first point: The findings could have been compared more critically to findings of previous studies and the theoretical contribution could be highlighted more explicitly. The recommendations for future studies (e.g. which features are relevant for different emotions) were interesting, but it is not argued clearly what this would add to theoretical understanding.

We are glad that the reviewer found our discussion to cover interesting material. We have further added more elaborate explanations regarding the theoretical significance of our suggestions.

We have added two major points to our discussion, also inspired by the study by Pollux et al. 2014 which the reviewer kindly drew our attention to:

> *"[...]regarding the results of the present experiments, we would predict that the basic mechanisms outlined here will be replicated with other facial expressions, but perhaps not*

for the same features. For example, happy-surprised pairings should be best recognized from the mouth, as both involve mainly muscles of the lower face half [10,13,32]. Interestingly, eye-tracking studies found that more fixations on the eye regions correlates with improved emotion recognition for all expressions [14,33] and argue that fixating the eyes reflects a holistic processing strategy. One could test this assumption using happy-surprise pairings and the composite face task from experiment 2. If holistic processing is associated with fixating the eyes, then eye judgements should be easier in a composite face task, even if a strongly diagnostic distracting lower face half is present."

(Discussion Section, 8th paragraph)

And we have further clarified why the use of a discrimination task is warranted in future studies investigating the phenomenon:

"The employed identification task could be complemented by a discrimination task. There, it should be easier to discriminate two images when they are on opposing sides of the category boundary, compared with being both sampled from within the same category [5,6]. Thus, the sigmoid curves from the present experiments could be used to predict the discrimination performance in a follow-up experiment. As differences in slope between conditions can reflect differences in variance of the underlying noise distribution or the uncertainty of the observer [34,35], it would be important to establish that a between-category advantage [5] is indeed present at the point where the psychometric function is steepest in the identification task."

(Discussion Section, 9th paragraph)

Discussion of training facial expression categorization and the role of gaze focus could have been linked to a few existing studies examining related topics (e.g. Dadds, 2012; Pollux et al,

2014). More thought could be given to structure of the whole discussion section.

Thank you for the helpful references. In addition to our discussion of the Pollux et al. 2014 paper which we reference in the comment above, we now also discuss this study and its high relevance in our discussion of face recognition training:

> *"[...] Another interesting question is to what degree the failure to concentrate on one half of a composite face can be voluntarily overridden, especially since previous work has shown that emotion recognition improvements are associated with increased fixation of the eyes [33]. This could be investigated by introducing training blocks before the experiment proper, a forced minimum exposure time to discourage spontaneous responses, biofeedback from an eye-tracking device helping to fixate on a specific face part, and rewards for correct responses."*

(Discussion Section, 2nd paragraph before last)