New perspective on face learning: Stability modulates resolution of facial representations in the optimal observer

Morgan Reedy and Christel Devue*

School of Psychology, Victoria University of Wellington, New Zealand

*Corresponding author: Christel Devue, School of Psychology, Victoria University of Wellington, PO Box 600, Wellington 6040, New Zealand

Phone: +64(0)4 463 5898

Email: christel.devue@vuw.ac.nz

Word count:

Abstract: 148

Intro, discussion and authors' notes: 2063

Methods and results: 2127

References: 34

Abstract

Learning new faces is challenging, error prone, and subject to large individual differences. A dominant theory claims that we remember faces by creating averaged representations that discard irrelevant changes due to viewing circumstances and retain stable inner features (e.g., eyes). If so, why do we occasionally fail to recognise familiar people and why don’t we always encode inner features? We propose a parsimonious face encoding system, in which the relative stability of (extra)facial feature determines the resolution at which they are encoded, following a coarse-to-fine strategy. This is confirmed in an ecological learning paradigm where, all else being equal, faces with stable appearances were encoded more coarsely than faces with variable appearances. The framework assumes that individual differences are based on the ability to use cost-efficient encoding strategies flexibly. Accordingly, poor recognisers rigidly encode coarse features whereas a lack of stability encourages good recognisers to refine their representations.

Keywords: Face recognition, Individual differences, Face processing, Stability, Appearance, Parsimony

New perspective on face learning: Stability reduces resolution of facial representations in the optimal observer

The average person knows about 5000 faces (Jenkins, Dowsett, & Burton, 2018), a feat that even automatic algorithms come nowhere close to mimicking (Mandal, 2016). However, people's abilities vary (Bobak, Pampoulov, & Bate, 2016) and recognition of familiar faces can also be thrown off by simple changes in appearance (Carbon, 2008; Devue, Wride, & Grimshaw, 2018; Sinha & Poggio, 1996), which existing theories struggle to explain.

Theories based on computer modelling claim we encode idiosyncratic properties of familiar faces by averaging stable inner features (eyes, nose, mouth) and ignoring changeable aspects like hair or changes resulting from lighting (Burton, Jenkins, Hancock, & White, 2005; Jenkins & Burton, 2011; Kramer, Young, & Burton, 2018). While recognition of familiar people is usually robust, we struggle to recognise people we've just met because we rely on extra-facial features (Bruce, Henderson, Newman, & Burton, 2001; White, Kemp, Jenkins, Matheson, & Burton, 2014). Why don't we always rely on inner features if this is more efficient? And why do we occasionally fail to recognise familiar people after they change hairstyle or shave (Carbon, 2008; Devue et al., 2018; Sinha & Poggio, 1996)?

We are stuck with these contradictions because of strong assumptions that we rigidly encode the same group of inner features for all familiar faces and that peripheral features do not count. Consequently, numerous studies indiscriminately used images without peripheral features, yielding results that may not reflect real-world face learning. Recent research introduced ecological learning conditions using natural images and showed that exposure to variations (e.g., in lighting, viewpoint, expression, or appearance) during learning helps build reliable representations (Baker, Laurence, & Mondloch, 2017; Menon, Kemp, & White, 2018; Murphy, Ipser, Gaigg, & Cook, 2015; Ritchie & Burton, 2017). However, variations in natural images are often confounded and so the mechanism behind benefits variations provide remains undefined.

We propose a new parsimonious mechanism of face learning and familiarisation that reconciles past findings. We assume different facial information may be more or less diagnostic and relevant to memorise in individual faces, regardless of familiarity. Rather than always privileging inner features, features' diagnosticity is assessed based on the *stability* they exhibit over time (e.g., stable jawline compared to changing hairstyle). Stable features receive more representational weight than those who vary. Further, coarse features (head shape) are prioritised over fine details (nostril shape) to create cost-effective memory representations (Gao, Ding, Yang, Liang, & Shui, 2013) that refine over time, particularly if appearance changes or if demands for context-independent recognition increase. To allow recognition, invariant internal features (i.e., nose, mouth, eyes and cheekbones) of people who change their appearance more must receive more weight and be encoded in more detail (i.e., at higher resolution) than those of people with a consistent appearance. In these latter, extra-facial features are diagnostic and receive substantial representational weight, reducing the need for detailed encoding of inner features.

We developed video stimuli to test this framework in ecological learning conditions. Videos allow viewers to develop three-dimensional representations of to-be-learned faces (Baker et al., 2017; Pike, Kemp, Towell, & Phillips, 1997; Pilz, Thornton, & Bulthoff, 2006) and prevent simplistic encoding of image properties (Megreya & Burton, 2006). Participants studied 12 identities, half that displayed low and half high levels of variability in appearance (e.g., different hairstyle, makeup or facial hair), with each identity appearing in each condition for different participants. Importantly, the two sets of videos of each identity incorporated comparable levels of variations in factors known to facilitate learning (lighting, viewpoint, expression, or background), so as to isolate the specific contribution of variations in appearance over and beyond these other factors. We examined recognition and matching performance. To determine if stability in appearance affects the resolution at which stable inner features are encoded, test images only showed inner features. We expected that learning faces with high levels of variations in appearance would improve performance compared to learning the same faces with a consistent appearance. By contrast, theories that assume invariant inner aspects of individual faces are averaged across exemplars do not predict different averaging processes for inner aspects of faces whose appearance varies in different degrees. Therefore, they would predict similar levels of recognition in the two conditions.

Large individual differences exist across face processing tasks (e.g., Mccaffery, Robertson, Young, & Burton, 2018) and so an essential validation step of our framework is to test its ability to predict individual performance. Poor and good recognisers seemingly use different processing strategies: the latter treat internal aspects as wholes while poor recognisers use part-based analysis (Richler, Cheung, & Gauthier, 2011; Wang, Li, Fang, Tian, & Liu, 2012) and rely on extra-facial features (Duchaine & Nakayama, 2004; Murray, Hills, Bennetts, & Bate, 2018). Presumably, current theories would predict that good recognisers form robust representations of invariant internal features regardless of how much changeable aspects vary, which should result in similarly high performance with faces learned under low or high levels of variation. Since, on average, exposure to various types of variations improves recognition performance (Murphy et al., 2015), we may expect it will benefit others, perhaps poorer recognisers (despite an overall lower performance than good recognisers), by helping them realise that extra-facial features are undiagnostic and invariant aspects must be attended. In this scenario, one predicts a negative correlation between overall recognition performance and the extent to which learning benefits from increased levels of variations in appearance.

In our framework however, one's recognition skills lay in the ability to *distribute representational weights flexibly* to encode facial information *parsimoniously,* rather than in the exhaustive and detailed encoding of a fixed group of invariant features. If so, good recognisers should assess features' stability during learning and refine their representations of inner features more when variations in extra-facial features signal those are undiagnostic. Therefore, we expected increased performance for highly variable faces compared to less variable ones in good recognisers. By contrast, if poor recognisers *rigidly* focus on peripheral aspects that are not necessarily diagnostic (e.g., hairstyle), increased levels of variations should be confusing and reduce their ability to learn variable faces compared to faces with consistent appearance. We thus expected a positive correlation between overall recognition performance and the extent to which learning improves following exposure to increased variations.

Methods

Our design, analysis plan and predictions were pre-registered on Open Science Framework before data collection [https://osf.io/qajtr/register/5771ca429ad5a1020de2872e].

Participants. Based on power analyses (see Supplementary materials), we recruited 250 first year psychology students (197 women, 49 men, and 4 non-binary; Mean age = 19.3 $\pm$ 3.4 years, range 18-55). They gave informed consent and received course credits. We excluded 29 participants because response times were too fast in either task (mean RT below 500ms or 2.5 SD below the sample's mean RT in the corresponding task) and/or they failed 2 or more of 4 attention checks, leaving 221 participants (175 female, 43 male, and 3 non-binary; Mean age = 19.4 $\pm$ 3.5 years). The study was approved by the Human Ethics Committee of the School of Psychology.

Materials. In the learning phase, we used videos from online vlog channels devoted to book, movie or products reviews. They showed single individuals, indoors, talking to the camera in front of a limited set of different backgrounds and included changes in viewpoint (i.e., frontal to angled or profile view). We selected 12 foreign Vloggers (6 female and 6 male), unfamiliar to our participants, with large collections of videos and who had a consistent appearance across several videos but also showed notable changes in appearance in others. For each individual, ten clips with a consistent appearance (i.e., consistent hairstyle, makeup or facial hair) were created from different videos and used in the low variability condition. Moreover, two of these clips and 8 additional clips showing the individual with various changes in appearance (e.g., hair length, style or colour, makeup, facial hair, glasses) were used in the high variability condition. We visually matched the amount of changes in background and lighting present across videos between both conditions so that only levels of variation in appearance differed between an individual's two sets. Two versions of the learning phase were created in which the 6 identities (3 male, 3 female) learned under low variability or high variability were counterbalanced across participants.

Test images were created from 216 screenshots of the learned individuals from different videos than those used in the learning phase, to ensure recognition of identity rather than recognition of exemplars from specific videos. The same images were used at test regardless of whether a target identity had been learned under low or high variability condition. Images showed frontal or slightly angled views. Facial expressivity and gesture of individual vloggers were consistent in their different videos and so test images showed expressions (i.e., neutral, smiling, or resulting from speech) similar to those viewed in the two learning conditions. Six images per identity were used in the recognition task to provide several opportunities to recognise each identity. Twelve additional images were used in the matching task (8 on match trials, 4 on mismatch trials). An additional 216 screenshots of 36 different unlearned vloggers were selected in the same manner (i.e., matching in viewing angles and expressions). Twelve vloggers were used in the recognition task (6 images each). The 24 others were used in the matching task: 12 as similar looking foils for learned identities on mismatch trials (4 images each) and 12 in the unlearned condition (6 used as unlearned identities, 12 images each, and 6 used as foils for unlearned identities on mismatch trials, 4 images each).

We measured individual face recognition skills with the standard Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). Participants study 6 faces with 3 different viewpoints. At test, participants select learned faces amongst three choices (two foils and one learned identity). Over 72 trials, test pictures show increasing changes (lighting, inclusion/exclusion of external features, orientation, and image degradations).

Procedure. Participants completed the experiment on their personal computer through the testable (testable.org) platform. They proceeded through four parts in a fixed order: the CFMT, the learning phase of our paradigm, the recognition task, and the matching task. A link to our learning paradigm, which contains copyrighted material, is available on demand from the corresponding author.

Learning phase. The learning phase consisted of 12 randomised blocks, one per identity, that each contained ten 4-second video clips in a random order. Participants were explicitly told that each block showed one single individual and were asked to memorize the 12 faces. The goal of having participants learn such a high number of faces was to avoid ceiling effects and to be able to examine the impact of learning condition even in the best recognisers. To ensure participants were attending to the faces, they rated how agreeable, attractive, and memorable they found each person via 7-point Likert scales (1 = “Not at all” and 7 = “Very”) at the end of each block.

Recognition task. There were 144 randomized trials, 72 with learned identities (6 per identity) and 72 with unlearned faces (6 female and 6 male, with 6 trials each). Participants judged if they had learned the face or not.

Matching task. There were 144 randomised match and mismatch trials. Match trials presented pairs of images of the same individual. There were 48 match trials with learned faces (4 per identity), 24 trials in the high and 24 in the low variability condition. There were 24 unlearned match trials with 6 new identities not used in the recognition phase (4 trials each). Mismatch trials presented pairs of images including a learned face (24 trials in the high and 24 in the low variability condition, 4 trials per identity) or an unlearned face (24 trials, 4 per the same 6 identities as on match trials), paired with a foil with a similar physical appearance. Participants judged if the two images showed the same or different people.

Pre-planned analyses and predictions

We used signal detection to assess how exposure to variations affects sensitivity (*d’*), namely here, the ability to discriminate between inner features of learned and unlearned faces. For the sake of space and clarity, other pre-registered analyses and their results, that are redundant (i.e., on accuracy and reaction times) or less directly relevant (i.e., analyses of ratings performed during learning, analyses of associations between other measures in the two tasks) are presented in Supplementary Materials.

Recognition. Here, *d′* is calculated based on hit rate (correct recognition of learned faces) in each learning condition and overall false alarm rate (incorrect judgements of familiarity on unlearned faces). We compared *d′* in the high and low variability conditions with a paired samples t-test. We expected exposure to high levels of variation in appearance would lead to higher sensitivity to internal features than low levels of variation.

Matching. Here, *d′* incorporates match and mismatch trials, and reflects sensitivity to differences in identities (see Kingdom & Prins, 2016). Hits are correct mismatch trials (i.e., accurate judgment that two images show different identities). False alarms are errors on match trials (i.e., inaccurate judgment that pairs of images show different identities). We conducted a one-way repeated measures ANOVA with learning condition (high, low, unlearned) as within-subject factors on *d′*, followed up by three paired sample t-tests comparing high, low and unlearned conditions. We expected that *d′* may be higher for faces learned via high levels of variability compared to faces in the low variability or unlearned conditions.

Individual differences. The extent to which individual participants were affected by exposure to variability was measured through the *Variability Effect Index* (VEI). VEI was calculated in each task by subtracting *d*′ in the low variability condition from *d′* in the high variability condition. Note that a different formulation of VEI was pre-registered [(d′ High) – (d′ Low)/(d′ High) + (d′ Low)]. This formula was discarded because a few participants had higher false alarm rates than hit rates, leading to overall negative d′ which, unlike the new formulation, did not accurately reflect the impact of variability on individual performance. With the new calculation, positive VEIs indicate increased sensitivity to inner features from learning highly variable faces. Negative VEIs indicate decreased performance following exposure to increased variations in appearance compared to learning faces with more consistent appearances.

Associations with the CFMT. We calculated Pearson's correlations between scores on the standard Cambridge Face Memory Test (Duchaine & Nakayama, 2006) and VEI in each task. In the recognition task, we expected that good recognisers (i.e., people with higher scores) would benefit from high levels of variation more than poor recognisers who would suffer from it and rely on stability. So we expected a positive correlation between CFMT scores and VEI, and assessed this association with a one-tailed test. We had no specific expectation in the matching task and used a two-tailed test.

Associations within tasks. We examined correlations between overall accuracy and VEI in each task. Again, in the recognition task we expected a positive correlation between recognition accuracy and VEI and so we tested this association with a one-tailed test.

Associations between tasks. We examined associations between VEI in the recognition task and VEI in the matching task to test whether the impact of exposure to variations affects the memory and perceptual domains in the same direction.

Results

Recognition. As expected, on average, participants were significantly more sensitive to inner features of faces learned in the high variability condition (Mean d' = .972, SD = .736; 95% CI [0.875, 1.07]) compared to the low variability condition (Mean d' = .899, SD = .665; 95% CI [0.811, 0.987]), $t(220) = 2.22$, $p = .027$, $d = .149$ (95% CI for d [0.017, 0.282]); see **Figure 1** (top panel).

Matching. Overall, sensitivity on the matching task showed a significant main effect of learning condition, $F(2, 440) = 77.6$, $p < .001$, $\eta^2 = .261$; see **Figure 1** (bottom panel). Participants were more sensitive to differences in identity after learning highly variable faces (Mean d' = 2.39, SD = .671; 95% CI [2.3, 2.48]) compared to unlearned faces (Mean d' = 1.90, SD = .753; 95% CI [1.8, 1.99]), t(220) = 11.48, *p* < .001, *d* = .773 (95% CI for *d* [0.622, 0.922]), and for faces in the low variability condition (Mean d' = 2.34, SD = .653; 95% CI [2.25, 2.43]) compared to unlearned faces, t(220) = 10.02, *p* < .001, *d* = .674 (95% CI for *d* [0.527, 0.82]). Sensitivity in the high and low condition was not significantly different, t(220) = 1.10, *p* = .272, *d* = .074 (95% CI for *d* [-0.058, 0.206]).

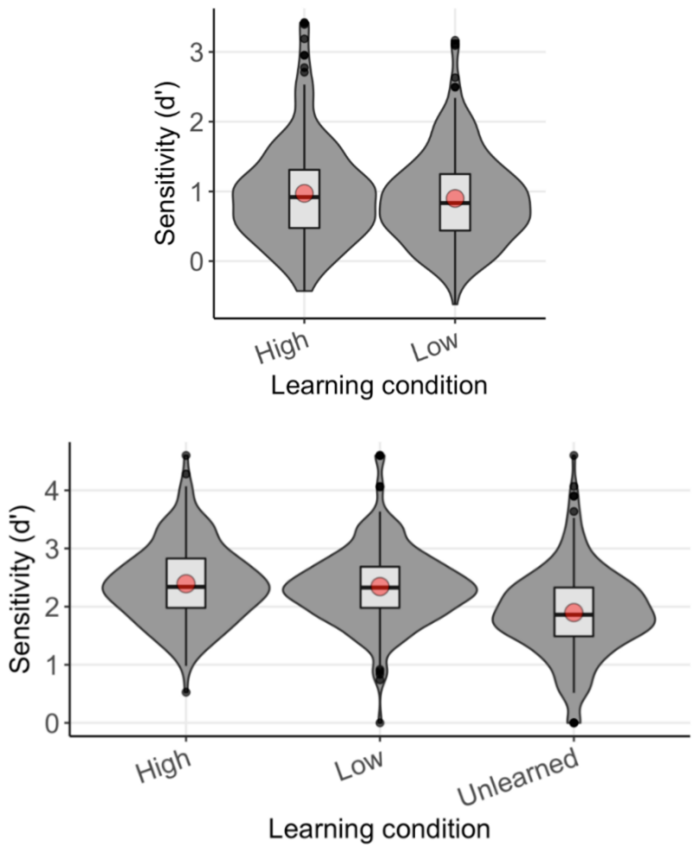

Figure 1. Sensitivity (*d'*) in the recognition task (top) and in the matching task (bottom). Mean values are represented by the red circles, boxplots show the distribution in quartiles, while the violins' width represents performance distributions across participants.

Individual Differences.

Associations with the CFMT. While CFMT scores predicted most aspects of performance (i.e., accuracy, d', hit rate, false alarm rate; see Supplementary materials), they did not significantly predict the extent to which exposure to high levels of variability (as indexed by VEIs) impacts individual performance in either task, $r_{\text{rec, one-tailed}}$ = .084, p = .107 (95% CI [-0.049, 0.214]); r_{match} = -.008, p = .900 (95% CI [-0.14, 0.124]), suggesting that the CFMT may not tap into that specific aspect of face processing. This is perhaps because in the CFMT, faces are learned in conditions showing limited variations (i.e., viewpoint only) or the response format (i.e., 3-alternative forced choice) allows using different strategies (e.g., elimination procedure) than in our two tasks.

Associations within tasks. **Figure 2** shows large individual differences in the way variations impact performance, ranging from large impairments (negative VEIs) to large improvements (positive VEIs) in sensitivity.

We found that the more one responded to high levels of variations by refining their representation of inner features, the more accurate they were in the recognition task. And as expected, recognition VEI and overall recognition accuracy were positively correlated, r = .157, $p_{one\text{-}tailed}$ = .0097 (95% CI [0.026, 0.283]). By contrast, in the matching task, despite large individual differences in the impact of variations, there was no significant correlation between matching VEI and overall accuracy, r = .019, p = .778 (95% CI [-0.113, 0.151]).

Associations between tasks. Improved sensitivity in the recognition task following exposure to increased variations at learning was associated with improved sensitivity to differences in identities in the matching task. This is indicated by a significant positive relationship between recognition and matching VEIs, r = .156, p = .02 (95% CI [0.024, 0.282]).

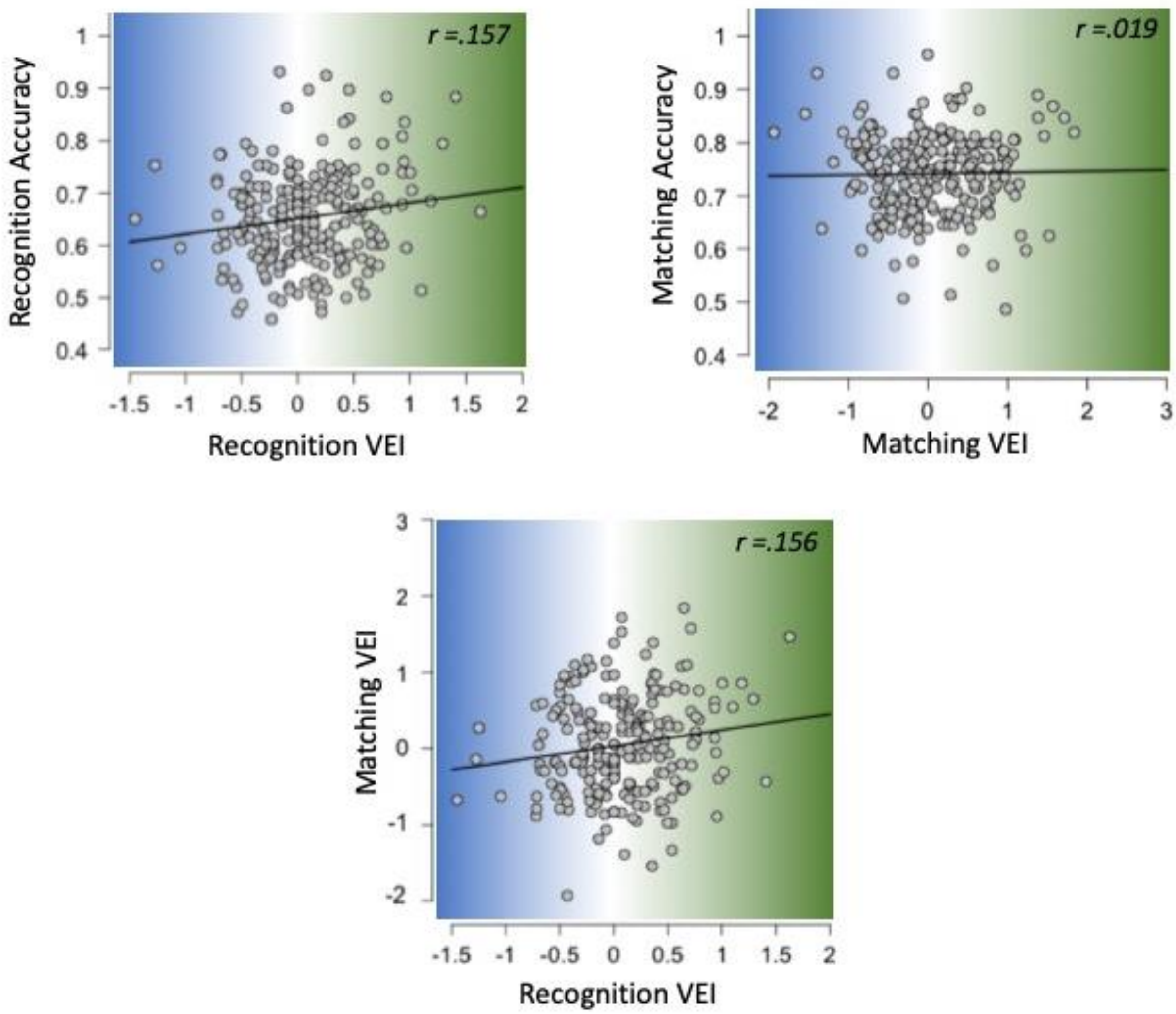

Figure 2. Associations between individual accuracy and individual impact of increased variation (indexed by VEI) in the recognition task (top left panel), in the matching task (top right panel), and association between VEIs in both tasks (bottom middle panel). Negative VEIs indicate an reduction in sensitivity following exposure to increased levels of variations during learning (low variability > high variability), positive VEIs indicate an improved sensitivity (low variability < high variability).

Discussion

Our new perspective on face familiarisation assumes a dynamic and parsimonious distribution of representational weights for individual faces, determined by the relative stability of changeable features, and operating in a coarse-to-fine direction. We used an ecological learning paradigm to test this rationale, in which changeable features of given individuals either remained stable or varied. We predicted that inner features of variable faces would be encoded in more details than faces with a consistent appearance in which coarse stable extra-facial features are also diagnostic.

We found that the focus on extra-facial features believed to characterise poor recognisers (Murray et al., 2018) and unfamiliar face processing (Johnston & Edmonds, 2009) occurs at least with some faces in most recognisers. As predicted, people's ability to recognise inner features depends on the relative stability that changeable extra-facial features exhibited during learning. When faces displayed large variations in appearance during learning, their inner features were better discriminated from inner features of strangers than when the same faces showed a consistent appearance.

For a parsimonious face recognition system, extra-facial features constitute useful identity cues and are deemed diagnostic when learning conditions suggest they will not change (and will be visible) in the future. Encoding coarse extra-facial features and facial configuration over inner facial details is cost-efficient because it incurs less storage demands (Goa et al., 2013). The relative stability of extra-facial and other changeable features (e.g., accessories) thus dictates the resolution at which invariant structural aspects are encoded to optimise recognition.

Although the impact of variations on recognition was small (Cohen's d = .149), our results are remarkable because faces in the two learning conditions were learned for the same duration and in similarly rich viewing conditions (i.e., including movement, changes in lighting, in background, and facial expression) known to help building rich three-dimensional representations

(Mileva et al., submitted; Pike et al., 1999; Pilz et al., 2006) and thought to lead to the creation of robust representations of inner facial features.

The benefit of rich viewing conditions is evident in the overall improvement in matching performance we found following learning (Clutterbuck & Johnston, 2002). Although levels of variations at learning did not affect performance in the matching task, participants who benefited from increased level of variations in one task also did in the other. Effects of increased variability are more modest on perceptual than on memory tasks (Robins et al., 2018), probably because there is no need for a prior detailed memory representation in a perceptual discrimination task that involves comparing two visible images. By contrast, a prior memory representation is essential in a recognition task that involves comparing one image to that representation.

Crucially, our framework also predicted individual recognition performance. Those that benefited from increased levels of variations in appearance were also the most accurate recognisers. Because test images only included inner features, the fact that good recognisers performed better in the high than in the low variability condition demonstrates that they relied on encoding stable extra-facial features in the low variability condition. If superior face recognition abilities resulted from a default indiscriminate encoding of all facial details, recognition performance of good recognisers should not have differed between the two learning conditions. Consistently, recent studies showed that both good and poor recognisers are affected by unexpected changes in extra-facial features between learning and test, even with highly familiar faces learned in naturalistic conditions (Devue et al., 2018; Esins, et al., 2016).

What distinguishes good from poor recognisers is their flexibility. It is optimal to be able to rely on stability that sustains the test of time, but to also be able to flexibly switch one's focus towards inner features and encode them in more details when extra-facial features prove unreliable. The least accurate recognisers were the most impaired by high variations in appearance, and so they were those who performed better after being exposed to stability, perhaps because stability gave them an opportunity to encode coarse internal configuration too. Unlike good recognisers that use a coarse encoding strategy only when it is cost-efficient, poor recognisers seem stuck with that strategy. The positive correlation between the impact of exposure to variations in the matching and in the recognition tasks also points to individual differences in the ability to make good use of variability in appearance. People differ vastly in their ability to assess stability, and those who are able to exploit variations to fine-tune their representations of someone's inner facial features show the most improvements in tasks that involve memory and perceptual discrimination.

We believe that our framework opens up many new research avenues, in which dichotomies that proved unhelpful in the past, between familiar and unfamiliar faces, or between poor and good recognisers, will become less relevant. Our framework challenges perennial ideas and resolves old controversies. For example, the idea that unfamiliar faces are processed based on pictorial encoding distinct from face-related encoding. If a parsimonious face encoding system assumes stability by default, it should not bet on change without signal to do so. When we learn new faces from single or limited sets of images, encoding coarse peripheral features or even pictorial artefacts is cost-efficient as long as they are diagnostic, that is, sufficient to distinguish different faces in the set. This results in cheap low resolution representations but leads to poor performance if unexpected change in viewing conditions or in appearance occur subsequently. Our framework also explains surprising effects with familiar faces and the anecdotal observation that swapping Al Gore's inner features for Bill Clinton's can go unnoticed (Sinha & Poggio, 1996). We hope that a more careful consideration of the characteristics of specific individual faces we deal with, as well as of the parameters of learning and testing conditions, along with individual differences in abilities to handle these parameters should help advance the field immensely.

Authors' notes.

We thank Sofie de Sena for her help with picture stimuli preparation and the Susilo lab for their support. A preliminary version of this manuscript was posted on the Open Science Framework website in February 2019 [https://osf.io/6s8c7]. The preregistration is visible at [https://osf.io/qajtr/register/5771ca429ad5a1020de2872e]. Aggregated de-identified data are visible at [https://osf.io/nve3g/]. Portions of the data were presented at KiwiCAM 2018 and at the 46th Experimental Psychology Conference.

References

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19–30. http://doi.org/10.1016/j.cognition.2016.12.012

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting Superior Face Recognition Skills in a Large Sample of Young British Adults. *Frontiers in Psychology*, *7:1378*. http://doi.org/10.3389/fpsyg.2016.01378

Brédart, S., & Young, A. W. (2004). Self-recognition in everyday life. *Cognitive Neuropsychiatry*, *9*(3), 183–97. http://doi.org/10.1080/13546800344000075

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207–218. http://doi.org/10.1037/1076-898X.7.3.207

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, *23*(1), 1–31. http://doi.org/10.1016/S0364-0213(99)80050-0

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: the power of averages. *Cognitive Psychology*, *51*(3), 256–84. http://doi.org/10.1016/j.cogpsych.2005.06.003

Carbon, C. C. (2008). Famous faces as icons. The illusion of being an expert in the recognition of famous faces. *Perception*, *37*(5), 801–806. http://doi.org/10.1068/p5789

Clutterbuck, R., & Johnston, R. a. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985–994. http://doi.org/10.1068/p3335

Devue, C., Wride, A., & Grimshaw, G. M. (2018). New insights on real-world human face recognition. *Journal of Experimental Psychology: General*. Advanced online publication. http://doi.org/10.1037/xge0000493

Duchaine, B. C., & Nakayama, K. (2004). Developmental prosopagnosia and the Benton Facial Recognition Test. *Neurology*, *62*(7), 1219–1220. http://doi.org/10.1212/01.WNL.0000118297.03161.B3

Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bülthoff, I. (2016). Face perception and test reliabilities in congenital prosopagnosia in seven tests. *I-Perception*, *7*(1). http://doi.org/10.1177/2041669515625797

Gao, Z., Ding, X., Yang, T., Liang, J., & Shui, R. (2013). Coarse-to-fine construction for high-resolution representation in visual working memory. *PLoS One, 8(2)*:e57913.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, *62*(2), 201–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9141907

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1571), 1671–1683. http://doi.org/10.1098/rstb.2010.0379

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society B: Biological Sciences*, *285*. http://doi.org/10.1098/rspb.2018.1319

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: a review. *Memory*, *17*(5), 577–596. http://doi.org/10.1080/09658210902976969

Kingdom, F., & Prins, N. (2016). Psychophysics: A practical introduction (2nd edition). Academic Press, London.

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*(June 2017), 46–58. http://doi.org/10.1016/j.cognition.2017.12.005

Longmore, C. A., Santos, I. M., Silva, C. F., Hall, A., Faloyin, D., & Little, E. (2017). Image dependency in the recognition of newly learnt faces. *Quarterly Journal of Experimental Psychology*, *70*(5), 863–873. http://doi.org/10.1080/17470218.2016.1236825

Mandal, B. (2016). Face Recognition: Perspectives from the Real-World. Retrieved May 8, 2018, from http://arxiv.org/abs/1602.02999

Mccaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(21), 1–15. http://doi.org/10.1186/s41235-018-0112-9

Megreya, A. M., & Burton, a M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition*, *34*(4), 865–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17063917

Menon, N., Kemp, R. I., & White, D. (2018). More than a sum of parts : robust face recognition by integrating variation. *Royal Society Open Science*, *5*, 172381.

Mileva, V.R., Bobak. A.K, Stewart, C., Cougan, L., & Hancock, P.J.B. (submitted). Learning faces from variability using video-clips.

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 577–581. http://doi.org/10.1037/xhp0000049

Murray, E., Hills, P. J., Bennetts, R. J., & Bate, S. (2018). Identifying Hallmark Symptoms of Developmental Prosopagnosia for Non-Experts. *Scientific Reports*, *8*(1), 1–12. http://doi.org/10.1038/s41598-018-20089-7

Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing Moving Faces: The Relative Contribution of Motion and Perspective View Information. *Visual Cognition*, *4*(4), 409–438. http://doi.org/10.1080/713756769

Pilz, K. S., Thornton, I. M., & Bulthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research*, *171*(4), 436–447. http://doi.org/10.1007/s00221-005-0283-8

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464–71. http://doi.org/10.1177/0956797611401753

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *The Quarterly Journal of Experimental Psychology*, *70*(5), 1–9. http://doi.org/10.1080/17470218.2015.1136656

Robins, E., Susilo, T., Ritchie, K., & Devue, C. (2018). Within-person variability promotes learning of internal facial features and facilitates perceptual discrimination and memory. Retrieved from https://osf.io/8tndq/

Sinha, P., & Poggio, T. (1996). I think I know that face... *Nature*, *384*(6608), 404–404. http://doi.org/10.1038/384404a0

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. http://doi.org/10.1177/0956797611420575

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, a. M. (2014). Passport Officers' Errors in Face Matching. *PLoS ONE*, *9*(8), e103510. http://doi.org/10.1371/journal.pone.0103510