Preprint
Vol. 0 (2023), pp.~1–22

ISSN XXX-XXXX-XXXX

Multidimensional Signal Detection Modeling Reveals Gestalt-Like Perceptual Integration of Face Emotion and Identity

S. Sanaz Hosseini & Fabian A. Soto
Florida International University
Department of Psychology

Numerous studies have tested the hypothesis that facial identity and emotional expression are independently processed, but a solid conclusion has been difficult to reach, with the literature showing contradictory results. We argue that this is partly due to different researchers using different definitions of perceptual integration and independence, usually vague and/or simply operational, and also due to lack of proper stimulus control. Here, we performed a study using 3-D realistic computer-generated faces for which the discriminability of identities and expressions, the intensity of the expressions, and low-level features of the faces were controlled. A large number of participants, distributed across twelve experimental groups, performed identification tasks for the six basic emotional expressions and the neutral expression (between 2018 and 2019). A multidimensional signal detection model was utilized to analyze the data, which allowed us to distinguish between multiple formally-defined notions of independence and holism. Results showed strong and robust violations of perceptual independence that were consistent across all experiments and suggest Gestalt-like perceptual integration of face identity and expression. To date, our results provide the strongest evidence for holistic/Gestalt processing found among face perception studies that have used formal definitions of independence and holism.

Keywords: Face perception, face identity, face emotion, perceptual integration, perceptual independence, signal detection theory, general recognition theory

The human face can convey important information about multiple social and psychological cues and states, including emotional expression and multiple aspects of face identity (e.g., race, sex, age). In vision science, it has been proposed that such features are processed independently or separately (e.g., Bernstein & Yovel, 2015; Bruce & Young, 1986; Duchaine & Yovel, 2015; Haxby et al., 2000). In contrast, the affective science literature suggests the opposite, with situational context being important for inferences of emotion from facial expression (Aviezer et al., 2017; Hess & Hareli, 2019), either to clarify the ambiguous mapping between expressions and the emotions eliciting them (Barrett et al., 2019) and/or to quickly and easily obtain socially-relevant information from faces (Adams et al., 2017). Part of this situational context is the individual presenting the expression, and a body of literature shows that variables related to identity influence emotion perception (Albohn & Adams, 2016; Albohn et al., 2019). This suggests that not only are identity features integrated into holistic percepts (Piepers & Robbins, 2012), but such percepts might also integrate emotional expression information.

Results of direct tests of the independence hypothesis are contradictory (e.g., D'Argembeau et al., 2003; Etcoff, 1984; Campbell, 1996; Campbell & Burke, 2009; Pell & Richards, 2013; Fox et al., 2008; Schweinberger & Soukup, 1998; Fitousi & Wenger, 2013), likely because researchers have used a variety of definitions of perceptual integration and independence, most of them vague and simply operational (see Mestry et al., 2012; Richler et al., 2012). This issue is compounded if one takes into account the emotion literature where, due the multimodal nature of emotion perception (Keltner et al., 2019), operational tests might measure not only a variety of visual processes, but other perceptual and post-perceptual processes as well. Indeed, theories positing independent pathways for identity and expression would explain observations of integration (i.e., context-specific processing) of these face dimensions as post-perceptual effects (or pre-perceptual effects in the stimuli, see below).

A way to bring clarity to this literature is to formalize concepts like integration and independence using mathematical theories (e.g., O'Toole et al., 2001). When such formal definitions are developed and linked to specific behavioral tasks, it is clear that the exact same dimensions can be interactive according to one definition, but independent according to a different definition

(Richler et al., 2012). General recognition theory (GRT) is an extension of signal detection theory developed to study interactions between stimulus dimensions (Ashby & Townsend, 1986; Ashby & Soto, 2015). Figure 1 shows the GRT model for an experiment testing independence of face identity and emotional expression, where stimuli are combinations of two identities and two expressions. The contours represent two-dimensional Gaussian distributions, each summarizing the noisy perceptual representation of a single stimulus. The straight lines are decision bounds that divide the space into four response regions, one for the identification of each stimulus.

GRT defines three types of independence between dimensions, with violations of each type representing a different form of integrated processing. *Perceptual separability* is the type of independence that research on visual perception of face identity and expression has aimed to test (e.g., through the Garner interference task; e.g., Fitousi & Wenger 2013; Wang et al. 2013). In perceptually separable dimensions, changes in the stimulus along one dimension do not alter the perceptual representation of the stimulus on the other dimension. In Figure 1, identity is perceptually separable from expression, as indicated by identical overlapping marginal distributions (the unidimensional Gaussians shown on each dimension) for each identity regardless of expression. However, expression is not perceptually separable from identity, as Bob's expressions are perceived as angrier than Joe's expressions. This formal definition of perceptual separability captures nicely the vernacular notion of "invariance" used in vision science (e.g., Anzellotti & Caramazza, 2014).

Perceptual independence is defined for the representation of each individual stimulus (i.e., each ellipse), and it holds when noise along one dimension is independent of noise along the other dimension. In Figure 1, the representation of angry Joe shows a violation of perceptual independence, illustrated as a tilted ellipse, which reveals that perception of the face as belonging to Joe is related to perception of that face as being angry (i.e., correlated perceptual noise). This captures the vernacular notion of a Gestalt or holistic perception (Townsend & Wenger, 2014), as in this case the perception of two aspects of a face (e.g., anger and identity) are inextricably linked. A review of the literature reveals consistent lack of evidence for holistic perception of face features defined in this way (for a review, see Townsend & Wenger, 2014).

Finally, *decisional separability* holds when the decision bound for classification along one dimension does not change with changes in the level of the other dimension (for a definition in terms of piecewise linear bounds, see Ashby & Lee, 1991). This form of independence is represented by bounds that are orthogonal to the dimension that they classify; any deviation from orthogonality is a violation of decisional separability. For instance, figure 1 shows that decisional separability is violated for identity but it holds for expression. Post-perceptual forms of interaction are captured by this concept.

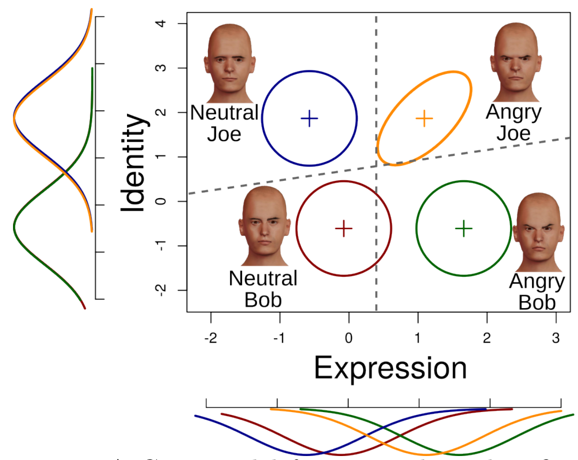


Figure 1. A GRT model for a complete identification task. Each dimension has two levels, resulting in four stimuli that are perceptually represented by bivariate Gaussian distributions. Ellipses represent contours of equal likelihood from such distribution. Marginal distributions are shown on each dimension. The dotted lines represent decision bounds that partition the space into four response regions.

Only some combinations of tasks and analyses are able to differentiate the three forms of independence, and most research on the interaction between face identity and emotional expression has not achieve that differentiation. For example, the Garner interference task (Garner, 1974) measures a combination of perceptual and decisional separability (Ashby & Maddox, 1994). Here, we use a complete identification task (Ashby & Soto, 2015; Soto et al., 2017) and a model-based analysis using GRT-wIND (GRT with individual differences; Soto et al., 2015). This approach can distinguish between all three forms of independence, at the cost of an assumption that the underlying pattern of the perception of a set of stimuli is similar among people, and differences in performance arise from attentional and decisional strategies that they employ in a task (for a discussion, see Silbert & Thomas, 2013, 2017; Soto et al., 2015).

A second issue likely to contribute to contradictory results in the literature is lack of stimulus control, as the results of many studies could partially mirror dependencies found in the experimental stimuli rather than reflecting any underlying perceptual processes (Tjan & Legge, 1998). This issue is key if one wants to suggest that a given effect is related to the way in which emotional expression is perceived rather than produced. To reach such conclusions, stimuli should be controlled in at least three ways. First, the relevant stimulus dimensions should not be correlated with any irrelevant dimension. Low-level features of the face (e.g., head shape, hairstyle, eye color, skin tone and texture) should be controlled because, without such control, perceptual separability could be trivially achieved through attention to salient low-level features (Anzellotti & Caramazza, 2014). Second, the intensity of the relevant dimensions should be constant across stimuli. For example, if person A is more expressive in demonstrating emotions than person B, then an observer may find discriminating emotions for person B harder compared to person A (a violation of perceptual separability), due to stimulus properties rather than perceptual processes. Third, the dimensions should be equally discriminable, as discriminability might reduce interference (Ganel & Goshen-Gottstein, 2004; Wang et al., 2013).

Here, we achieved these three goals by (1) using highly controlled identity 3D models, designed to be realistic while eliminating low-level confounds, (2) creating realistic models of expressions that can be applied over all of the identities to equalize the displayed emotion across them, and (3) equating discriminability of the identities and expressions based on the results of a pilot study. In addition, we determine the independence of face identity and expression separately for each of Ekman's six basic emotions: anger, disgust, fear, happiness, sadness and surprise (Ekman & Friesen, 1975). Most previous studies have targeted a subset of these expressions and, in most cases, participants were asked to discriminate a specific pair of expressions. In our study, we evaluate each individual expression on its own, discriminated against a neutral expression. This makes results comparable across expressions, allowing us to evaluate if independence is found for some expressions and not others, and it is closer to the task that people encounter during naturalistic face perception. Finally, we used two different sets of identities per expression to assess the generalizability of the outcome across faces, for a total of twelve experiments[1].

Methods

Participants

Four hundred and twenty undergraduate students from Florida International University were recruited through SONA. Of those who reported demogrphics, 28.57% were male and 71.42% were female, with ages from 18 to 39 years of age (mean = 21.21). We aimed to obtain data from 30 participants in each group. After exclusion of uninformative datasets (see below), we obtained between 20 and 30 participants in each experiment. Simulation work has shown that these sample sizes provide high parameter recoverability for the model used in our analyses (Soto et al., 2021). Participation was voluntary and compensated with course credit. The Institutional Review Board of Florida International University approved this study, and written informed consent was obtained from all participants. Data were collected in 2018 and 2019.

Stimuli

Highly controlled three-dimensional identity models and expression pose models were generated using computer graphics software (MakeHuman, www.makehumancommunity.org; see Hays et al., 2020) so that the exact same expression model could be applied on different identity models. Four identity models were grouped into sets of two, with each pair controlled to have identical facial hair, head size, head shape, hairstyle, hair color, eye color, skin color, skin tone, skin texture, and facial fat. In addition, they were all the same age, had the same level of sex (maleness), and race (all caucasian). Alongside the identity models, six expression models for the emotions happy, sad, anger, fear, surprise, and disgust were created, all inspired by a real actor's poses (validated in Hays et al., 2020). An expression model can be applied on any identity model, which eliminates any difference in the intensity of the displayed emotion by different identities. To equate the discriminability of the dimensions, a pilot study was conducted to determine differences between levels of each dimension (e.g., neutral vs. sad) discriminated at average sensitivity of $d' = 1.5$ (Macmillan & Creelman, 2005). Details and results of the pilot study, and renders of the models can be found in the *Supplementary Material* (*SM*).

Procedures

Twelve experiments were performed, each using a different stimulus set, which resulted from the combination of six basic expressions and two identity sets. Each participant was presented with only one stimulus set.

To avoid image-matching strategies (Burton & Jenkins, 2011), participants were familiarized with the identities in the set by watching short movies of each face accompanied by its name (Bob or Joe). Participants were instructed to memorize the faces and informed that they

[1]We refer to them as experiments rather than experimental groups, as they were performed sequentially within a time window of sixteen months, without random assignment of participants.

would be later tested on their recognition (as in Megreya & Burton, 2006).

After familiarization, participants performed a complete identification task, in which they were asked to identify the specific combination of identity and expression presented by pressing a key. ("B", "G", "J", and "N"; see Figure S4 and description in *SM*).

Each trial started with a 500ms fixation cross, followed by a 200ms stimulus. Participants could press a key within 2s after stimulus onset; after this the message "Too slow!" was displayed in red. When a response was recorded, feedback ("Correct!" in blue or "Incorrect!" in red) was displayed for 1s. Feedback for slow or incorrect responses was followed by a 5s penalty timeout, to motivate participants to perform accurately.

Participants completed twenty-three blocks of trials. Each block included five repetitions of each stimulus, randomized within the block, yielding a total of 460 trials.

Model-Based Data Analysis

The following steps were taken to perform separate model-based analyses of all twelve datasets.

The data from participants whose performance was lower than 40% (near chance) or higher than 90% (near perfect) correct was excluded, as they provide little to no information for model-fitting. For each participant, we excluded data acquired during learning of the task, following the procedures outlined in Soto et al. (2015) (for details on included/excluded participants and data, see Table S2 in *SM*). Included data were aggregated into a 4×4 confusion matrix where each row corresponded to one stimulus and each column to one response key. The cells of this matrix contained the frequency of each response when the participant was presented with each of the stimuli.

We used the package *grtools* (Soto et al., 2017) to fit the GRT-wIND model (Soto et al., 2015) to the confusion matrices using maximum likelihood estimation. This model assumes that some properties of the perceptual distributions are the same across participants (i.e., perceptual separability and independence hold or fail for all participants). Individual differences in performance are captured by differences in decisional and attentional processes. To ensure finding the global maximum of the likelihood function, the optimization algorithm was run 120 times with random starting values, and the GRT-wIND model with the highest maximum likelihood was kept.

We performed likelihood ratio tests of perceptual separability, perceptual independence, and decisional separability. Each test compares the fit of the full estimated model with that of a model restricted to show a form of independence, determining whether or not significant violations in that form of independence were observed. All the tests were applied as implemented in *grtools*. Effect size equations are not available for these tests, but in the *SM* (Figures S5-S7) we report deviations from perceptual separability and independence that can be interpreted as unstandardized effect sizes, together with their 95% confidence intervals.

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and research materials are available at `https,//osf.io/thjkq/?view_only=e17041f72471468993a0444d968c1e27`. Data were analyzed using R, version 3.6.2 (R Core Team, 2021) and the package *grtools*, version 0.3.1 (Soto et al., 2017). This study was not pre-registered.

Results

The best-fitting GRT models from each experiment are shown in 2. The model achieved good fits to the data, accounting for between 89.63% and 97.95% of the variance (see Table S3 in *SM*). Each graph in Figure 2 follows the same structure as Figure 1 and should be interpreted in the same way. In addition, the results of likelihood ratio tests have been summarized in Table 1, where results that are consistent across stimulus sets are highlighted in bold and italics.

The most consistent and strongest effect, found across all experiments, was that identity and expression are not perceptually independent. Figure 2 shows that the pattern of violations of perceptual independence was the same across all experiments, a "flower" pattern in which stronger perceptual evidence for the correct level of one dimension was accompanied by stronger evidence for the correct level of the other dimension. Note that, because the midpoint of each axis represents weaker perceptual evidence in the corresponding dimension, the perceptual evidence for both stimulus dimensions is positively correlated across all stimuli.

The results of likelihood ratio tests for perceptual independence supported this conclusion, with all tests showing strong and significant violations of perceptual independence: anger - set 1, $\chi^2(4) = 6477.9$, $p < 0.001$, anger - set 2, , $\chi^2(4) = 6110.19$, $p < 0.001$, disgust - set 1, $\chi^2(4) = 3771.7$, $p < 0.001$, disgust - set 2, $\chi^2(4) = 6777.1$, $p < 0.001$, fear - set 1, $\chi^2(4) = 5292.53$, $p < 0.001$, fear - set 2, $\chi^2(4) = 3974$, $p < 0.001$, happiness - set 1, $\chi^2(4) = 5990.1$, $p < 0.001$, happiness - set 2, $\chi^2(4) = 5211.3$, $p < 0.001$, sadness - set 1, $\chi^2(4) = 5382.9$, $p < 0.001$, sadness - set 2, $\chi^2(4) = 4798.3$, $p < 0.001$, surprise - set 1, $\chi^2(4) = 6504.9$, $p < 0.001$, surprise - set 2, $\chi^2(4) = 8754.6$, $p < 0.001$.

The second most consistent effect was that of violations of decisional separability, with 21 out of 24 tests showing significant violations of decisional separability. Results were inconsistent across stimulus sets only for the expressions of disgust and sadness (see Table 1). Regarding the decisional separability of identity, we found significant results for anger - set 1, $\chi^2(25) = 132.2$, $p < 0.001$, anger - set 2, $\chi^2(27) = 237.96$, $p < 0.001$, disgust - set 1, $\chi^2(22) = 56.2$, $p < 0.001$, fear - set 1, $\chi^2(21) = 123.37$, $p < 0.001$, fear - set 2, $\chi^2(22) = 300.9$, $p < 0.001$, happiness - set 1, $\chi^2(23) = 92.8$, $p < 0.001$, happiness - set 2, $\chi^2(28) = 138.9$, $p < 0.001$, sadness - set 1, $\chi^2(28) = 203.01$, $p < 0.001$, surprise - set 1, $\chi^2(24) = 119.7$, $p < 0.001$, surprise - set 2, $\chi^2(27) = 151.5$, $p < 0.001$. Only the tests for disgust - set 2 ($p > 0.1$) and sadness - set 2 ($p = 0.051$) did not reach significance, with the latter being quite close and perhaps reflecting low statistical power. Regarding the decisional separability of expression, we found significant results for anger - set 1, $\chi^2(25) = 128.6$, $p < 0.001$, anger - set 2, $\chi^2(27) = 267.32$, $p < 0.001$, disgust - set 1, $\chi^2(22) = 80.8$, $p < 0.001$, fear - set 1, $\chi^2(21) = 105.69$, $p < 0.001$, fear - set 2, $\chi^2(22) = 297.9$, $p < 0.001$, happiness - set 1, $\chi^2(23) = 126.4$, $p < 0.001$, happiness - set 2, $\chi^2(28) = 384.9$, $p < 0.001$, sadness - set 1, $\chi^2(28) = 292.19$, $p < 0.001$, sadness - set 2, $\chi^2(23) = 70.6$, $p < 0.001$, surprise - set 1, $\chi^2(24) = 176.3$, $p < 0.001$, surprise - set 2, $\chi^2(27) = 125$, $p < 0.001$. Only the test for disgust - set 2 did not reach significance ($p > 0.1$).

Finally, perceptual separability showed results that were in general stimulus-specific; that is, whether or not violations of perceptual separability were observed was highly dependent on the stimulus set (i.e., specific identities) with which participants were tested. There were two exceptions in which results were consistent across stimulus sets. First, expression of anger was consistently separable from identity and vice-versa (all $p > 0.1$). Second, there was an asymmetric pattern of perceptual separability observed for happy expressions, expression was perceptually separable from identity (all $p > 0.05$), but identity was not perceptually separable from expression (set 1: $\chi^2(2) = 119.3$, $p < 0.001$; set 2: $\chi^2(2) = 199.1$, $p < 0.001$).

For other expressions, we found significant violations of perceptual separability for one stimulus set, but not the other. For disgust, we found significant results for expression ($\chi^2(2) = 36.4$, $p < 0.001$) and identity ($\chi^2(2) = 14$, $p = 0.001$) in stimulus set 1, but neither was significant in stimulus set 2 ($p > 0.1$). For fear, we found significant results for expression ($\chi^2(2) = 16.9$, $p < 0.001$) and identity ($\chi^2(2) = 65.8$, $p < 0.001$) in stimulus set 2, but neither was significant in stimulus set 1 ($p > 0.5$). For sadness, we found significant results for expression ($\chi^2(2) = 139.3$, $p < 0.001$) and identity ($\chi^2(23) = 14.2$, $p < 0.001$) in stimulus set 2, but neither was significant in stimulus set 1 ($p > 0.05$). Finally, for surprise, we found significant results for expression ($\chi^2(2) = 99$, $p < 0.001$) and identity ($\chi^2(2) = 20.1$, $p < 0.001$) in stimulus set 2, but neither was significant in stimulus set 1 ($p > 0.05$).

We performed a post-hoc control experiment with morphed identity dimensions, which are known to be highly interactive according to multiple tests (e.g., Blunden et al., 2015; Soto & Ashby, 2015). This provided a benchmark of what magnitude of violations of perceptual independence and separability to expect from interactive dimensions using our procedures. Detailed results from this study are reported in the *SM*, and they all were in line with the results of our main analysis with likelihood ratio tests. When compared to our benchmark, the magnitude of violations of perceptual separability was highly variable and strongly dependent on stimulus set. On the other hand, the magnitude of violations of perceptual independence were consistently larger than the benchmark. This suggests that our main finding of holistic processing of expression and identity was not a mere artifact of our methods or the type of stimuli that we used, a possibility suggested by the absence of such findings in the face perception literature (for a review, see Townsend & Wenger, 2014).

Discussion

We performed a study with the goal of determining whether facial identity and expression are perceptually independent, while controlling for a number of confounds in previous similar studies. We controlled stimulus confounds by creating three-dimensional models of identity and expression pose, and by precisely manipulating the discriminability of identities and expressions. We controlled decisional confounds by performing a GRT model-based analysis of data obtained from an identification task, which additionally allowed us to evaluate multiple formally-defined forms of independence, rather than relying on vague operational definitions. The study included two experiments for each of Ekman's six basic emotional expressions.

The most consistent and strongest result observed was a significant violation of perceptual independence of identity and expression, showing that when the expression of a face is perceived more distinctively, then its identity is also perceived more distinctively, and when the expression of a face is perceived with higher ambiguity, then there is higher ambiguity in its perceived identity. In other words, participants were unable to

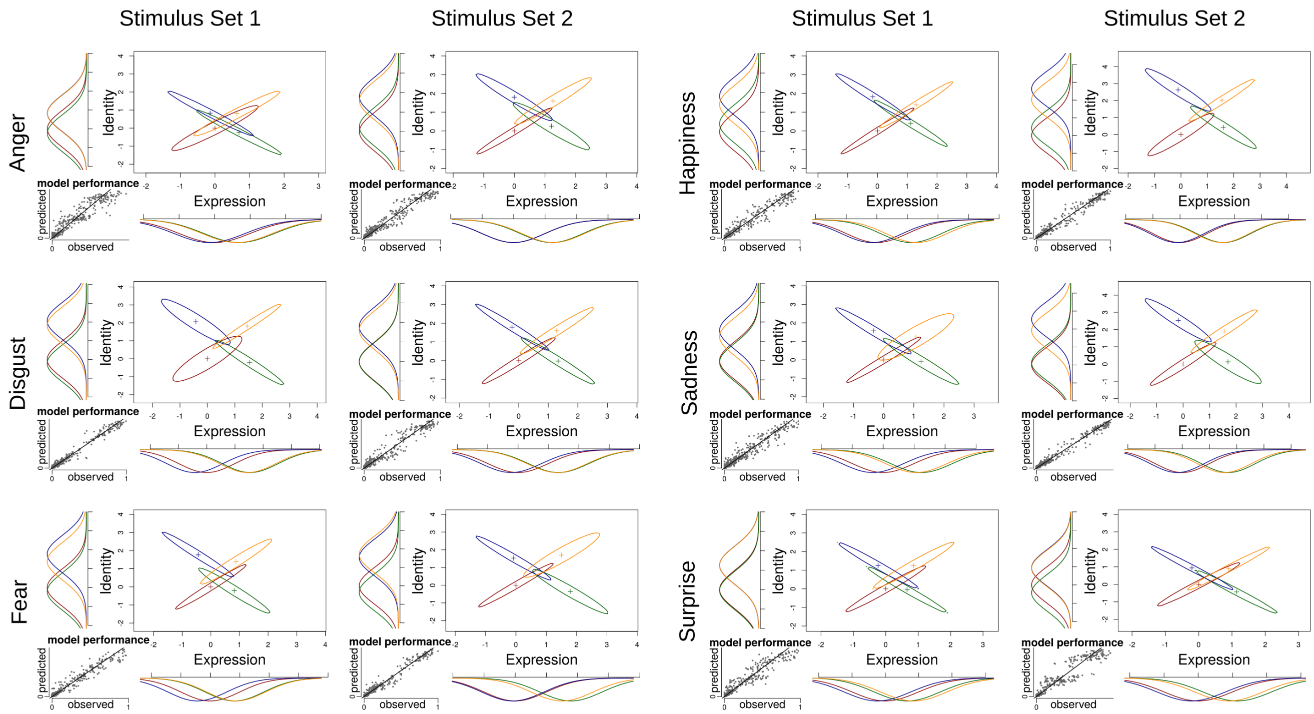


Figure 2. Best-fitting GRT model from each of the experiments in this study (one sub-figure per study). Within each sub-figure, Each contour of a different color corresponds to one of the stimuli included in the experiment (red, neutral Bob, green, expressive Bob, blue, neutral Joe, and orange, expressive Joe), and the marginal distributions for each of these contours are represented by univariate Gaussians with the same color, presented to the left of the y-axis for identity and below the x-axis for expression. The panel at the bottom-left of a sub-figure shows a comparison between the response probabilities predicted by the best-fitting model (y-axis) against the observed response proportions (x-axis) for all the participants. That is, for each stimulus in the task (e.g., "angry Joe") a participant could choose one of four different responses ("angry Joe", "neutral Joe", "angry Bob", or "neutral Bob"), with their proportion throughout the experiment adding up to one. The same data could be obtained from the model, and the scatterplot shows a comparison of such probabilities across all stimuli and participants.

Table 1

Summary of results of likelihood ratio tests in the main study. Results that are consistent across stimulus sets are highlighted using a font in bold and italics. PS = Perceptual Separability, PI = Perceptual Independence, DS = Decisional Separability.

Expression	PS (Expression)	PS (Identity)	PI	DS (Expression)	DS (Identity)
Anger / Stimulus Set 1	***Yes***	***Yes***	***No***	***No***	***No***
Anger / Stimulus Set 2	***Yes***	***Yes***	***No***	***No***	***No***
Disgust / Stimulus Set 1	No	No	***No***	No	No
Disgust / Stimulus Set 2	Yes	Yes	***No***	Yes	Yes
Fear / Stimulus Set 1	Yes	Yes	***No***	***No***	***No***
Fear / Stimulus Set 2	No	No	***No***	***No***	***No***
Happiness / Stimulus Set 1	***Yes***	***No***	***No***	***No***	***No***
Happiness / Stimulus Set 2	***Yes***	***No***	***No***	***No***	***No***
Sadness / Stimulus Set 1	Yes	Yes	***No***	***No***	No
Sadness / Stimulus Set 2	No	No	***No***	***No***	Yes
Surprise / Stimulus Set 1	Yes	Yes	***No***	***No***	***No***
Surprise / Stimulus Set 2	No	No	***No***	***No***	***No***

perceive a face's identity without its expression, or vice versa. This pattern strongly implies what other researchers (see Townsend & Wenger, 2014) have interpreted as holistic or Gestalt perception of face identity and expression. To the best of our knowledge, our results provide the strongest evidence for holistic/Gestalt processing found among face perception studies that have used formal definitions of holism (e.g., Cornes et al., 2011; Richler et al., 2008; for a review, see Townsend & Wenger, 2014). The pattern of violations of perceptual independence consistently found across all experiments was specific to the dimensions of identity and expression; it was not found in our benchmark study (see *SM*) or in other studies involving different facial dimensions (Martin et al., 2022).

This result was also absent from previous studies that evaluated independence of face identity and expression using GRT (Fitousi & Wenger, 2013; Soto et al., 2015). The most likely explanation for our novel result lies in the strong level of stimulus control used in our study, which allowed us to isolate the perceptual mechanisms involved in perception of face shape only. We hypothesize strongly integrated perception of changeable and stable aspects of face shape, against the predictions of traditional models of face processing (Bruce & Young, 1986; Haxby et al., 2000) and in line with more recent models of face perception that propose highly context-specific processing of emotional expression (Adams et al., 2017; Aviezer et al., 2017; Hess & Hareli, 2019), as well as those that identify a pathway in the visual system processing all aspects of face shape (Bernstein & Yovel, 2015; Duchaine & Yovel, 2015). Such integrated processing might be obscured by differences in non-shape features that facilitate the perception of stable aspects of face shape, such as identity, but not the perception of changeable aspects of face shape, such as expression.

The second most consistent result found in our study was that of violations of decisional separability, a common finding in the face perception literature (e.g., Fitousi & Wenger, 2013; Richler et al., 2008; Wenger & Ingvalson, 2002). This underscores the importance of using paradigms able to distinguish between decisional and perceptual forms of dimensional interaction. Some of the most popular tasks used in the literature on independence of face identity and expression, such as the Garner interference task (e.g., Etcoff, 1984; Fitousi & Wenger, 2013; Schweinberger & Soukup, 1998; Wang et al., 2013), are unable to distinguish between decisional and perceptual separability, and they provide no information about perceptual independence. It is important that future research avoids the use of such tasks, focusing instead on tasks and analyses that can distinguish between perceptual and decisional processes.

Finally, the observed pattern of results regarding perceptual separability was highly variable and dependent on both the specific expression and identities tested. This is again in line with recent research showing interactions between emotion perception and situational context variables, including identity (Albohn & Adams, 2016; Albohn et al., 2019). As suggested by an ecological approach to face perception (Adams et al., 2017), integration of information about identity and expression might be important because it allows observers to exploit any information available that is useful for social action and survival. This view predicts that integration should be stimulus-specific, as physiognomic features should influence potentially available information about emotion (e.g., different neutral faces would appear more or less expressive, see Albohn et al., 2019) and what identity variables could be inferred from expression.

Results were consistent across identities only for the expressions of anger and happiness. First, both anger and happiness showed perceptual separability from identity. Such consistency in the perception of a *level of expression* is different from advantages that these expressions might have for recognition or detection (Nummenmaa & Calvo, 2015). Happy and angry faces are interesting from an ecological perspective in that they are signals of safety and threat, respectively, and there is evidence suggesting that they evolved by exploiting already-existing perceptual mechanisms of survival (e.g., Franklin et al., 2019). Perhaps reliably evaluating the level of threat and safety in a face, regardless of who is showing these expressions, has a higher adaptive value than extracting more or less evidence of threat and safety depending on a person's physiognomy (e.g., dominance features, related to physical strength).

Second, happiness caused identities to be less perceptually distinctive, whereas anger had no effect. Regarding the effect of happiness, extraction of information about safety might be prioritized to such an extent that it always reduces processing of identity information. This is to be expected from prioritization of one source of information that is highly integrated with another. If that is the case, then it is hard to explain why identity perception was not influenced by changes in anger. One possibility is that threat processing is prioritized to such an extent that a parallel system has evolved to accomplish it. To explain violations of perceptual independence, however, that system would have to implement integration at some stage. More research is necessary to test this and other possibilities suggested by the current results.

We took advantage of a full factorial identification task design. This has several advantages over the famous Garner interference task (Garner, 1974), which measures a combination of perceptual separability and decisional separability (Ashby & Maddox, 1994). The inability of the task to distinguish perceptual from decisional factors could fairly explain why prior studies have resulted in contradictory conclusions (e.g., Etcoff, 1984; Ganel & Goshen-Gottstein, 2004; Schweinberger & Soukup, 1998; Stoesz & Jakobson, 2013). In addition, different studies have targeted different expressions, and we found great variation in violations of perceptual separability and decisional separability across different expressions. If the Garner task measures a form of separability not defined within GRT (Algom & Fitousi, 2016), then it remains to be determined exactly what that is (using a rigorous formal definition) and why it cannot be measured consistently across experiments. We believe that, without such advances, the Garner task is a poor choice to study perceptual independence/integration, and should be replaced by more precise approaches (e.g., Fitousi & Wenger, 2013; Townsend & Wenger, 2014; Fifić & Townsend, 2010).

Together, violations of perceptual independence and separability suggest that emotion perception shows context-specificity that is partially perceptual in nature, in line with the idea that facial communication exploits pre-existing perceptual mechanisms to generate socially-relevant integrated representations (Adams et al., 2017) and to disambiguate facial signals of emotion that vary widely across individuals (Barrett et al., 2019). Indeed, our results show that identity and expression cues are so tightly integrated in face perception that they act as a single perceptual unit, with both being either perceived or missed together in conditions of low signal.

Basic emotion theory (for a review, see Keltner et al., 2019) proposes that people reliably perceive emotion in faces showing a particular pattern of expression, and it is linked to the idea that the brain activity accompanying that perception is similarly stereotypical (see discussion in Lindquist et al., 2012). While context-specificity has been taken as evidence against basic emotion theory, the type of results that warrant such interpretation are different from those found here. Our results only show that identity can change *how strongly* an emotion is perceived, not *what* emotion is perceived. Similarly, our results support very specific hypotheses about how integration of expression and identity is achieved in the brain. Neurocomputational extensions of GRT (Soto et al., 2018) suggest that failures of perceptual separability and independence must be accompanied by similar context-specific neural encoding (i.e., failures of encoding separability and independence). Populations of neurons that encode expression and identity might be localized in specific brain areas or be more malleable (Lindquist et al., 2012), encapsulated in visual cortex or open to top-down influence (Firestone & Scholl, 2016). Regardless, neurocomputational theory indicates that they must be context-sensitive, in that changes in identity modify encoding of expression, and vice-versa.

Limitations

Our results might not generalize to people with demographics different from those of our participants. Although our stimuli were realistic and obtained from validated models (Hays et al., 2020), they do not match real-world face photographs. Our goal was to manipulate inner face shape features and standardize everything else about our models, and we believe that this level of control is behind the consistency of our results regarding perceptual independence. On the other hand, it is well known that people use other cues for face identification. Thus, our conclusions are limited to face shape perception, and not other aspects of face perception. Our model-based analysis depends on the assumption that patterns of interference between emotion and identity are similar across people, which has been called into question (Silbert & Thomas, 2017). Future research should explore potential individual differences and their implications for our conclusions, using traditional GRT analysis of a larger identification task (e.g., with three levels per dimension). Because model-fitting requires identification errors, we excluded data from a minority of participants with near-perfect performance, who might process faces differently from other people, showing higher levels of independence of processing. The definitions of independence and integration included in GRT refer to the nature of the perceptual representation of a stimulus, but other forms of independence can be formally defined which are beyond the scope of GRT. For example, systems factorial technology (see Little et al., 2017) offers definitions in terms of real-time processing, neurocomputational GRT (Soto et al., 2018) offers definitions in terms of neural representation, and extensions of GRT to the linear-nonlinear observer model (Soto, 2019) offer definitions in terms of information sampled by a perceptual observer (i.e., representations estimated in reverse correlation studies). We advice researchers to focus on such formal definitions and known ways to validly test them, rather than vague operational definitions.

Acknowledgements

This work was supported by the National Science Foundation under grant No. 2020982 to Fabian A. Soto. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Adams, R. B., Albohn, D. N., & Kveraga, K. (2017). Social vision: Applying a social-functional approach to face and expression perception. *Current Directions in Psychological Science*, 26(3), 243–248.

Albohn, D. N. & Adams, R. B. (2016). Social vision: At the intersection of vision and person perception. In J. R. Absher & J. Cloutier (Eds.), *Neuroimaging personality, social cognition, and character* (pp. 159–186). Elsevier.

Albohn, D. N., Brandenburg, J. C., & Adams, R. B. (2019). Perceiving emotion in the" neutral" face: A powerful mechanism of person perception. In U. Hess & S. Hareli (Eds.), *The Social Nature of Emotion Expression: What Emotions Can Tell Us About the World* (pp. 25–47). Springer International Publishing.

Algom, D. & Fitousi, D. (2016). Half a century of research on Garner interference and the separability–integrality distinction. *Psychological Bulletin*, 142(12), 1352–1383.

Anzellotti, S. & Caramazza, A. (2014). The neural mechanisms for the recognition of face identity in humans. *Frontiers in Psychology*, 5, 672.

Ashby, F. G. & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, 120(2), 150–172.

Ashby, F. G. & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, 38(4), 423–466.

Ashby, F. G. & Soto, F. A. (2015). Multidimensional signal detection theory. In *The Oxford handbook of computational and mathematical psychology*, Oxford library of psychology (pp. 13–34). New York, NY, US: Oxford University Press.

Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2), 154–179.

Aviezer, H., Ensenberg, N., & Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology*, 17, 47–54.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68.

Bernstein, B. & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience & Biobehavioral Reviews*, 55, 536–546.

Blunden, A. G., Wang, T., Griffiths, D. W., & Little, D. R. (2015). Logical-rules and the classification of integral dimensions: individual differences in the processing of arbitrary dimensions. *Frontiers in Psychology*, 5, 1531.

Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77(3), 305–327.

Burton, A. M. & Jenkins, R. (2011). Unfamiliar Face Perception. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), *Oxford Handbook of Face Perception* (pp. 287–306). Oxford University Press.

Campbell, J. & Burke, D. (2009). Evidence that identity-dependent and identity-independent neural populations are recruited in the perception of five basic emotional facial expressions. *Vision Research*, 49(12), 1532–1540.

Campbell, R. (1996). Dissociating face processing skills: decisions about lip read speech, expression, and identity. *The Quarterly Journal of Experimental Psychology: Section A*, 49(2), 295–314.

Cornes, K., Donnelly, N., Godwin, H., & Wenger, M. J. (2011). Perceptual and decisional factors influencing the discrimination of inversion in the Thatcher illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 645.

D'Argembeau, A., van der Linden, M., Comblain, C., & Etienne, A. M. (2003). The effects of happy and angry expressions on identity and expression memory for unfamiliar faces. *Cognition and Emotion*, 17(4), 609–622.

Duchaine, B. & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, 1(1), 393–416.

Ekman, P. & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues.* Unmasking the face: A guide to recognizing emotions from facial clues. Oxford, England: Prentice-Hall.

Etcoff, N. L. (1984). Selective attention to facial identity and facial emotion. *Neuropsychologia*, 22(3), 281–295.

Fifić, M. & Townsend, J. T. (2010). Information-processing alternatives to holistic perception: Identifying the mechanisms of secondary-level holism within a categorization paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1290–1313.

Firestone, C. & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, 39, e229.

Fitousi, D. & Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 133–155.

Fox, C. J., Oruç, I., & Barton, J. J. S. (2008). It doesn't matter how you feel. The facial identity aftereffect is invariant to changes in facial expression. *Journal of Vision*, 8(3), 11–11.

Franklin, r. G., Adams, R. B., Steiner, T. G., & Zebrowitz, L. A. (2019). Reading the lines in the face: The contribution of angularity and roundness to perceptions of facial anger and joy. *Emotion*, 19(2), 209–218.

Ganel, T. & Goshen-Gottstein, Y. (2004). Effects of Familiarity on the Perceptual Integrality of the Identity and Expression of Faces: The Parallel-Route Hypothesis Revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 583–597.

Garner, W. R. (1974). *The processing of information and structure*. New York: Lawrence Erlbaum Associates.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–232.

Hays, J. S., Wong, C., & Soto, F. A. (2020). FaReT: A free and open-source toolkit of three-dimensional models and software to study face perception. *Behavior Research Methods*, 52(6), 2604–2622.

Hess, U. & Hareli, S. (2019). The emotion-based inferences in context (ebic) model. In U. Hess & S. Hareli (Eds.), *The Social Nature of Emotion Expression: What Emotions Can Tell Us About the World* (pp. 1–6). Springer International Publishing.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*, 43, 133–160.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and brain sciences*, 35(3), 121–143.

Little, D., Altieri, N., Fific, M., & Yang, C. T. (2017). *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms*. Academic Press.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates, 2nd edition.

Martin, E. R., Hays, J. S., & Soto, F. A. (2022). Face shape and motion are perceptually separable: Support for a revised model of face processing.

Megreya, A. M. & Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition*, 34(4), 865–876.

Mestry, N., Wenger, M. J., & Donnelly, N. (2012). Identifying sources of configurality in three face processing tasks. *Frontiers in Perception Science*, 3, 456.

Nummenmaa, L. & Calvo, m. G. (2015). Dissociation between recognition and detection advantage for facial expressions: a meta-analysis. *Emotion*, 15(2), 243–256.

O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 1–38). Lawrence Erlbaum Associates Publishers.

Pell, P. J. & Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Research*, 79, 1–7.

Piepers, D. & Robbins, R. (2012). A Review and Clarification of the Terms "holistic," "configural," and "relational" in the Face Perception Literature. *Frontiers in Psychology*, 3.

R Core Team (2021). R: A language and environment for statistical computing.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 34(2), 328–342.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Perception Science*, 3, 553.

Schweinberger, S. R. & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human perception and performance*, 24(6), 1748.

Silbert, N. H. & Thomas, R. D. (2013). Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review*, 20(1), 1–20.

Silbert, N. H. & Thomas, R. D. (2017). Identifiability and testability in grt with individual differences. *Journal of Mathematical Psychology*, 77, 187–196.

Soto, F. A. (2019). Linking general recognition theory and classification images to study invariance and configurality of visual representations. *Journal of Vision*, 19(10), 87d.

Soto, F. A. & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, 139, 105–129.

Soto, F. A., Stewart, R. A., Hosseini, S., Hays, J. S., & Beevers, C. G. (2021). A computational account of the mechanisms underlying face perception biases in depression. *Journal of Abnormal Psychology*.

Soto, F. A., Vucovich, L., Musgrave, R., & Ashby, F. G. (2015). General recognition theory with individual differences: a new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review*, 22(1), 88–111.

Soto, F. A., Vucovich, L. E., & Ashby, F. G. (2018). Linking signal detection theory and encoding models to reveal independent neural representations from neuroimaging data. *PLoS Computational Biology*, 14(10), e1006470.

Soto, F. A., Zheng, E., Fonseca, J., & Ashby, F. G. (2017). Testing separability and independence of perceptual dimensions with general recognition theory: A tutorial and new R package (grtools). *Frontiers in Psychology*, 8.

Stoesz, B. M. & Jakobson, L. S. (2013). A sex difference in interference between identity and expression judgments with static but not dynamic faces. *Journal of Vision*, 13(5), 26–26.

Tjan, B. S. & Legge, G. E. (1998). The viewpoint complexity of an object-recognition task. *Vision Research*, 38(15), 2335–2350.

Townsend, J. T. & Wenger, M. J. (2014). On the dynamic perceptual characteristics of Gestalten: Theory-based methods. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization Get access Arrow* (pp. 948–968). Oxford, UK: Oxford University Press.

Wang, Y., Fu, X., Johnston, R. A., & Yan, Z. (2013). Discriminability effect on Garner interference: evidence from recognition of facial identity and expression. *Frontiers in psychology*, 4, 943.

Wenger, M. J. & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5), 872.

Supplementary Material
Multidimensional Signal Detection Modeling Reveals Gestalt-Like Perceptual Integration of Face Emotion and Identity

S. Sanaz Hosseini & Fabian A. Soto

Contents

1 Pilot Study

1.1 Methods

Highly controlled three-dimensional identity models were created using the software MakeHuman, such that models of eyeballs, eyebrows, skin, and teeth were identical. Two pairs of male identities were chosen to be used in this study. Expression models were generated by selecting an actor from the KDE+ database (Lundqvist et al., 1998), so that his photographs served as the basis for the creation of pose models for all six basic emotional expressions (Ekman & Friesen, 1975): anger, happiness, sadness, fear, disgust, and surprise. All the expression models were applied to all four selected identity models to construct the final expressive facial models plus the neutral version of each identity, creating 28 different facial (identity/expression) models, which were rendered from a frontal viewpoint. Identities were grouped into two stimulus sets (top two identities and bottom two identities in Figure S3). The availability of two separate stimulus sets allowed to test the generalizability of results across identities in the main experiments.

Using the software JPsychMorph, we could make morphed series of these renders, which varied either across identities or expressions. Morphed series are sequences of images obtained by interpolation between two extremes. Those extremes were two renders from the already generated facial models, and the interpolations were performed in 10% steps. To obtain psychometric curves for identity, two sequences of static stimuli were created, one for each stimulus set (i.e., pair of identities), going from one identity to the other in the set in 10% steps. The expression was kept neutral. The resulting stimuli varied in identity but kept expression constant. To obtain psychometric curves for expression, six different sequences of stimuli were created for each stimulus set, going from neutral to each of the six emotional expressions in 10% steps. The average of the two identities in the set was used, so that the resulting stimuli varied in expression but kept identity constant.

Figure S1 demonstrates one of the morph series created for the pilot study. This sequence was created for the average identity of one of the sets from the neutral expression to the happy expression with 10% steps.

A discrimination task was developed where groups of 20 participants (undergraduate students from Florida International University recruited through SONA) were first trained to discriminate the two extremes of the sequence for a total of 48 trials. This was followed by 50 blocks of testing, each involving 6 training

Figure S1: A morphed sequence from neutral (left) to happy (right) expressions within an identity. The sequence was generated in 10% steps.

trials with the extremes of the sequence and 9 testing trials with all the faces between the two extremes. Stimuli were presented for 200 ms in each trial (as in the main study). Participants were excluded if their accuracy across training trials did not reach 75% correct. Psychometric curves could be obtained from the data collected during both the expression discrimination and identity discrimination tasks (as shown in Figure S2). The data from each participant was fitted to a model of the psychometric curve based on signal detection theory (Lesmes et al., 2015) using the R package *quickpsy* (Linares & López-Moliner, 2016), which allowed to obtain sensitivity functions linking d' with morph level. The group model of the psychometric curves (the thick black psychometric curve in Figure S2) enabled us to find the morph levels for each sequence corresponding to the a sensitivity value of $d' = 1.5$, meaning that the discriminability between the selected level of the morphing sequence and the start point of the sequences (the first identity or the neutral expression) was kept constant across dimensions and stimulus sets.

1.2 Results

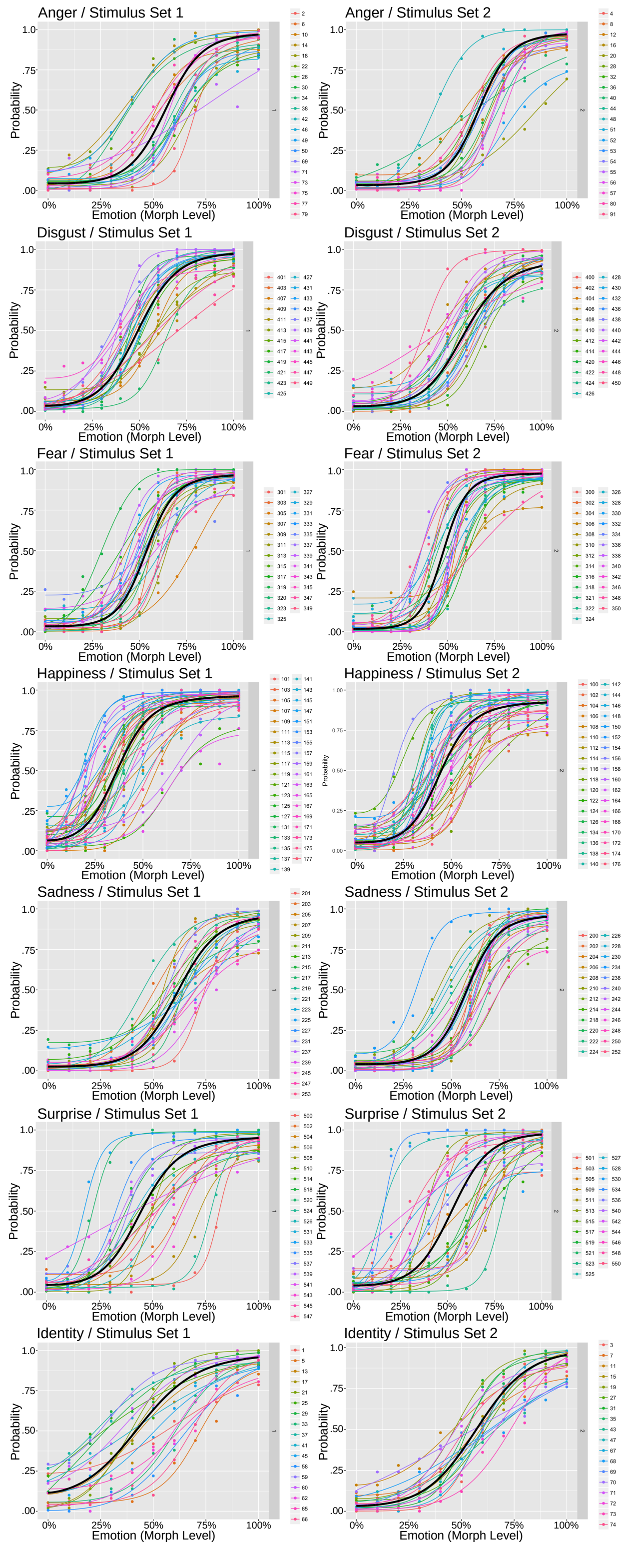


Figure S2: Psychometric curves obtained from the pilot study.

Figure S2 depicts the psychometric curve diagrams that we obtained for our pilot study. A psychometric curve visualizes the performance of the participants as a function of the stimulus level (the morph level in our study) in a discrimination task (e.g., identity 1 vs. identity 2, neutral vs. fearful, etc.). The data from each individual participant are shown in a different color, with points representing proportion of choices of the second of the two trained stimuli, and lines representing the best-fitting psychometric curve.

Depicted in black is the average psychometric curve for each group, obtained by averaging estimated parameters across participants. These group models were used to obtain thresholds corresponding to $d' = 1.5$, which are reported in Table S1. In the main experiment, we used morph levels equal to zero and the thresholds values reported in Table S1 as the two levels in each dimension, ensuring that the discriminability of dimensional levels was on average equivalent across stimulus sets.

Group	Threshold
Anger / Stimulus Set 1	52.51%
Anger / Stimulus Set 2	53.85%
Disgust / Stimulus Set 1	44.15%
Disgust / Stimulus Set 2	53.18%
Fear / Stimulus Set 1	49.50%
Fear / Stimulus Set 2	41.47%
Happiness / Stimulus Set 1	36.12%
Happiness / Stimulus Set 2	42.81%
Sadness / Stimulus Set 1	54.85%
Sadness / Stimulus Set 2	54.85%
Surprise / Stimulus Set 1	41.47%
Surprise / Stimulus Set 2	48.16%
Identity / Stimulus Set 1	49.16%
Identity / Stimulus Set 2	50.17%

Table S1: Thresholds obtained from the pilot experiment.

2 Main Study

2.1 Supplementary Methods

Renders from the full set of models used in our study are shown in Figure S3. Each row represents a different identity model, and each column represents a different expression pose model that could be applied independently to an identity model. Identity and expression are defined in MakeHuman as two separate sets of deviations from a base face model. Identity models are defined by a set of numbers quantifying the presence of specific anatomical face features (e.g., lip thickness, nose width, etc.). Expression pose models are defined by a set of numbers quantifying the presence of a particular expression pose (similar to an action unit: lip corner up, nose flare, etc.).

In our identification task, the keys "B", "G", "J", and "N" on a QWERTY keyboard were assigned to the stimuli, such that the keys on the left ("G" and "B") corresponded to Bob, the keys on the right ("J" and "N") correspond to Joe, the keys at the bottom ("B" and "N") corresponded to the neutral expression, and the keys at the top ("G" and "J") were assigned to the target expression This assignment of keys to stimuli was performed in order to highlight the factorial nature of the task (i.e., two simultaneous discriminations) and to help participants to more easily remember the keys. We asked participants to press the keys with the middle and index fingers of both hands, making the keys easier to work with (left hand would be assigned to Bob and right hand would be assigned to Joe, index fingers would report the neutral expressions and the middle fingers would report the target emotion expressions).

Figure S3: The identities and expression models used for the study. The identities in the first two rows are paired in the first set and the identities in the two last rows are paired in the second set of stimuli. The expressions are neutral, angry, happy, sad, fearful, disgusted, and surprised from left to right, respectively. Note that renderings of the original 3D models used to generate stimuli are shown to aid comparison; the final calibrated stimuli presented to participants showed less discriminable differences in expression and identity (see description in text).

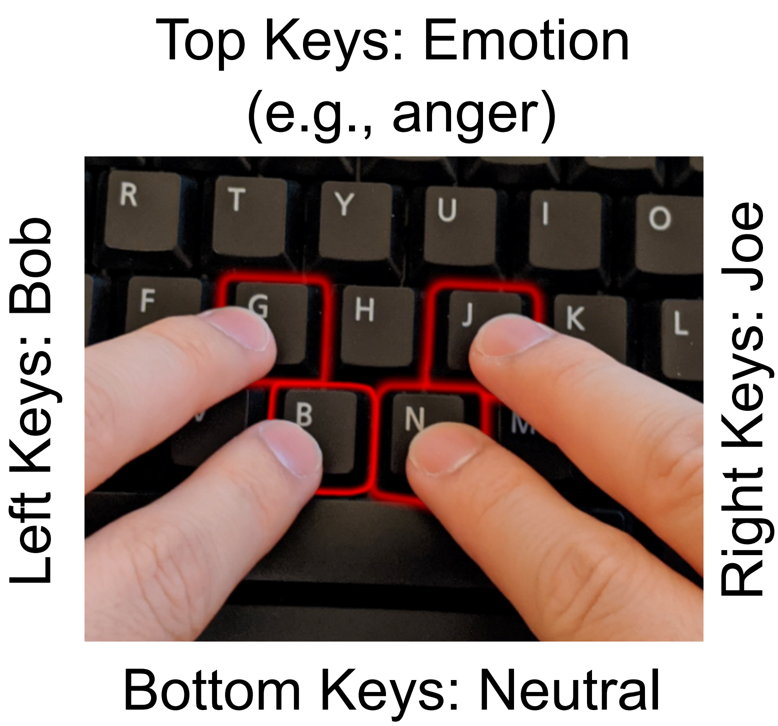


Figure S4: Participants were asked to use the response keys as demonstrated in this figure. The assignment of keys to stimuli shown in this figure was chosen to highlight the factorial nature of the task (i.e., two simultaneous discriminations) and to help participants to more easily remember the keys.

2.2 Supplementary Results

Group	Included	Excluded - Low Performance	Excluded - High Performance	Included trials (range, mean)	Excluded trials (range, mean)	Accuracy
Anger / Stimulus Set 1	25	4	1	216-460, 399	0-244, 62	60.22%
Anger / Stimulus Set 2	27	2	1	208-460, 386	0-253, 75	68.68%
Disgust / Stimulus Set 1	22	5	2	199-460, 385	0-261, 75	73.31%
Disgust / Stimulus Set 2	26	3	2	182-460, 435	0-278, 25	70.82%
Fear / Stimulus Set 1	21	8	1	234-460, 402	0-226, 58	64.59%
Fear / Stimulus Set 2	22	6	2	227-460, 414	0-233, 46	71.25%
Happiness / Stimulus Set 1	23	6	1	188-460, 384	0-272, 76	68.37%
Happiness / Stimulus Set 2	28	0	2	215-460, 423	0-245, 37	72.74%
Sadness / Stimulus Set 1	28	2	0	224-460, 424	0-236, 36	63.14%
Sadness / Stimulus Set 2	26	2	2	134-460, 412	0-326, 48	74.73%
Surprise / Stimulus Set 1	24	3	3	170-460, 417	0-290, 43	63.10%
Surprise / Stimulus Set 2	27	2	1	264-460, 443	0-196, 17	66.83%

Table S2: Number of included and excluded participants and trials, presented separately for each experiment.

Group	R^2
Anger / Stimulus Set 1	89.63%
Anger / Stimulus Set 2	94.47%
Disgust / Stimulus Set 1	97.36%
Disgust / Stimulus Set 2	94.45%
Fear / Stimulus Set 1	94.29%
Fear / Stimulus Set 2	96.96%
Happiness / Stimulus Set 1	94.22%
Happiness / Stimulus Set 2	96.93%
Sadness / Stimulus Set 1	92.95%
Sadness / Stimulus Set 2	97.95%
Surprise / Stimulus Set 1	91.04%
Surprise / Stimulus Set 2	90.63%

Table S3: Fit of the GRT-wIND model to data, presented separately for each experiment.

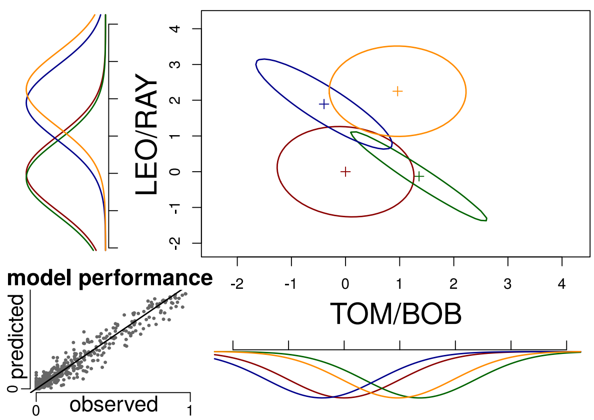
3 Benchmark Study

An issue with the results of likelihood ratio tests in our main study is that in many cases violations of perceptual separability were significant despite appearing to be weak in magnitude, compared to what we have seen in previous studies (e.g., Soto & Ashby, 2015). As we have argued elsewhere (Soto et al., 2017), such cases of weak but significant violations of perceptual separability are common and interpreting them is sometimes difficult without a benchmark of what one should expect from truly non-separable dimensions. To obtain such a benchmark, we performed a post-hoc control experiment with morphed identity dimensions, which are known to be highly interactive according to multiple tests of separability (Blunden et al., 2015; Folstein et al., 2012; Goldstone & Steyvers, 2001; Soto & Ashby, 2015). An additional advantage of this benchmark is that it allows us to determine to what extent the consistent and strong evidence of holistic (not perceptually independent) perception of face identity and expression observed in our experiments might be brought about by our methods and/or stimuli, as such evidence has been absent from the face perception literature (e.g., Cornes et al., 2011; Richler et al., 2008 ; for a review, see Townsend & Wenger, 2014).

Stimuli were obtained by morphing each of the identity models in set 1 with each of the identity models in set 2. This produced four morphed stimuli, which corresponded to the factorial combination of two levels of each identity dimension. A group of sixty participants completed an identification task involving those stimuli. All procedures were the same as in the main study, but participants were familiarized with all four original identities with the assigned names “Bob”, “Tom”, “Leo”, and “Ray”. After familiarization with the original identities, participants were presented with the four morphed identities (Tom/Leo, Tom/Ray, Bob/Leo, Bob/Ray) introduced as nephews of the original identities, to highlight that the morphed identities had combined features of those presented during familiarization. The initials of each name corresponded to the initials of the words “top”, “bottom”, “left,” and “right,” again to facilitate understanding and learning of the task.

Sixty participants completed the study. The data from 34 subjects were excluded from the final analysis due to performance being too high or low (see methods section of main manuscript). Data from 26 subjects in the dataset was used for the final analysis and the mean accuracy of the responses for these participants was at 61.56% correct.

The results obtained for the control dataset are reported in Figure 3. The model explains 93.44% of the variation in the data and the prediction errors made by the model do not follow a systematic pattern. The results suggest that there are violations of perceptual separability for both identity dimensions ($\chi^2(2) = 16.9$, $p < 0.001$ for dimension A and $\chi^2(2) = 25.8$, $p < 0.001$ for dimension B). Furthermore, decisional separability was violated for both identity dimensions ($\chi^2(26) = 247.5$, $p < 0.001$ for dimension A and $\chi^2(26) = 155.6$, $p < 0.001$ for dimension B). Violation of perceptual independence was also observed ($\chi^2(4) = 396.6$, $p < 0.001$).



Best-fitting GRT model from the control benchmark experiment. Ellipses are contours of equal likelihood, and associated marginal distributions are plotted with the same color. The sub-panel at the bottom-left shows the scatterplot of predicted response probabilities against observed response proportions. The diagonal line represents a model with perfect fit.

We used the obtained best-fitting models from all experiments in our main study and the control benchmark to estimate and compare the magnitude of deviations in perceptual separability and perceptual independence.

Deviations from perceptual separability were measured through the $L1$ distance between two distributions that share a value in a target dimension (e.g., emotional expression), but different values in an irrelevant dimension (e.g., identity),

$$L1 = \int |p_1(z) - p_2(z)|$$

where p_1 and p_2 represent the marginal distributions of values along the relevant dimension for two different levels (1 and 2) of the irrelevant dimension. To compute $L1$, we obtained 100 evenly spaced values of the dimensional variable z, starting five standard deviations to the left of the smaller of the two means of p_1 and p_2 , and ending five standard deviations to the right of the larger of those two means. Let z_k with k=1, 2, . . . 100 represent these discrete values of the relevant dimension z. An estimate of the of the summed L1 distances representing deviations from perceptual similarity was computed in the following way,

$$L1_d^G = \sum_m \sum_k |p_{m1}(z_k) - p_{m2}(z_k)|$$

Where d stands for the target dimension for which perceptual separability is being computed (either expression or identity), G stands for global as this is a measure that adds multiple $L1$ estimates (i.e., one for each level of the target dimension), and m indexes the levels of the target dimension.

Deviations from perceptual independence were measured as the sum of absolute values of the estimated correlation parameters.

To more easily compare the magnitude of deviations from perceptual separability and perceptual independence across groups, we obtained bootstrap confidence intervals. At each of 2,000 bootstrap steps, a simulated data set was sampled from the best-fitting GRT-wIND model, and was used to fit a new model to the simulated data. This model was in turn used to compute measures of violations of perceptual separability and independence as described above. For each measure, this results in an empirical distribution function of values, which was used to directly obtain 95% confidence intervals using a simple quantile procedure.

Figure S5 shows the obtained measures of deviations from perceptual separability for the expression dimension, together with their 95% bootstrap confidence intervals. The two leftmost plots represent the benchmark of perceptual separability violations provided by the control group, and the rest represent perceptual separability violations observed for all groups in the main study. Figure S6 shows the equivalent plots for deviations from perceptual separability for the identity dimension. In both cases, it can be seen that the results confirm what was observed from likelihood ratio tests, the magnitude of violations of perceptual separability is highly variable and in many cases strongly dependent on stimulus set. More importantly, such violations were for the most part comparable to those observed in the control experiment, although more so for the identity dimensions than for the expression dimensions.

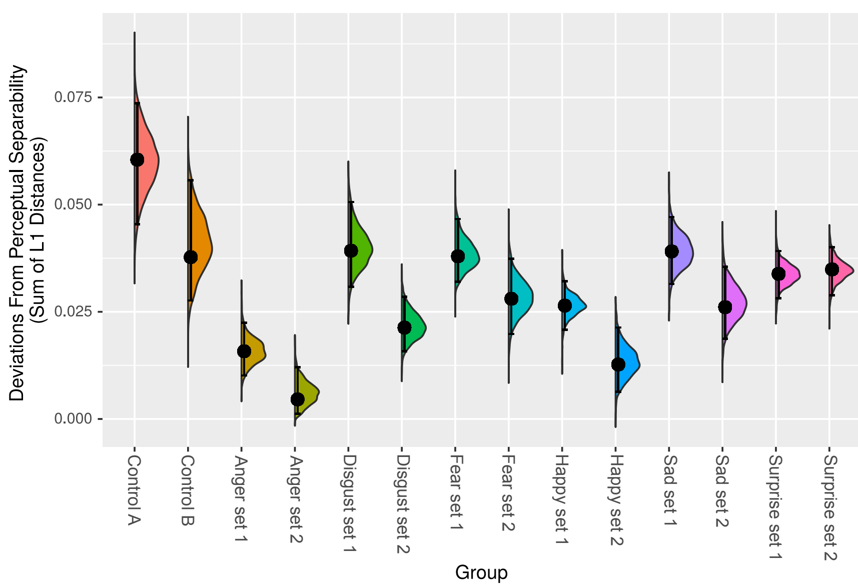


Figure S5: Measures of deviations from perceptual separability for the expression dimension, together with their 95% bootstrap confidence intervals. The two leftmost plots represent the benchmark of perceptual separability violations provided by the control group, and the rest represent perceptual separability violations observed for all experimental groups.

Figure S7 shows the obtained measures of deviations from perceptual independence, together with their 95% bootstrap confidence intervals. The leftmost plot represent the benchmark of perceptual independence violations provided by the control group, and the rest represent perceptual independence violations observed

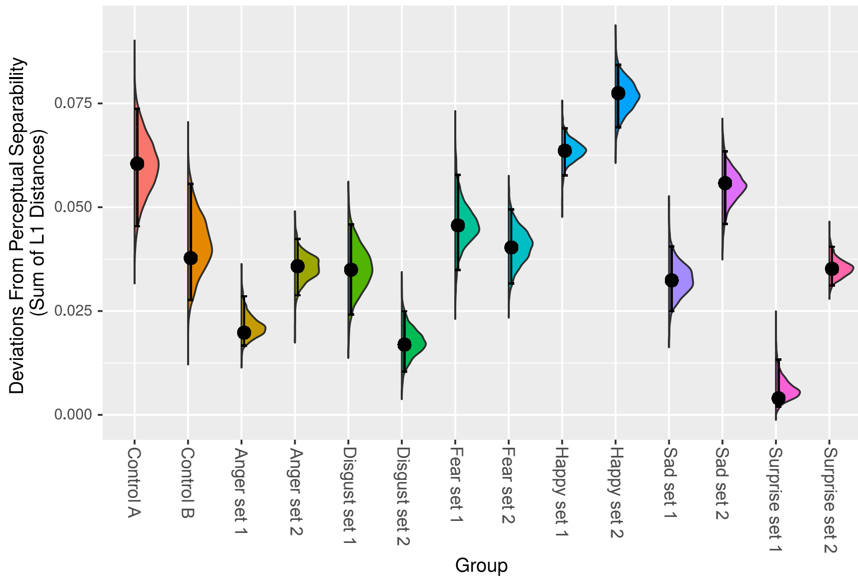


Figure S6: Measures of deviations from perceptual separability for the identity dimension, together with their 95% bootstrap confidence intervals. The two leftmost plots represent the benchmark of perceptual separability violations provided by the control group, and the rest represent perceptual separability violations observed for all experimental groups.

for all experimental groups. The results confirm that violations of perceptual independence in the main study were consistently larger than what was observed in the control group. This suggests that our main finding of holistic processing of expression and identity was not brought about by our methods and/or stimuli, as the same method and type of stimulus was used in the control experiment.

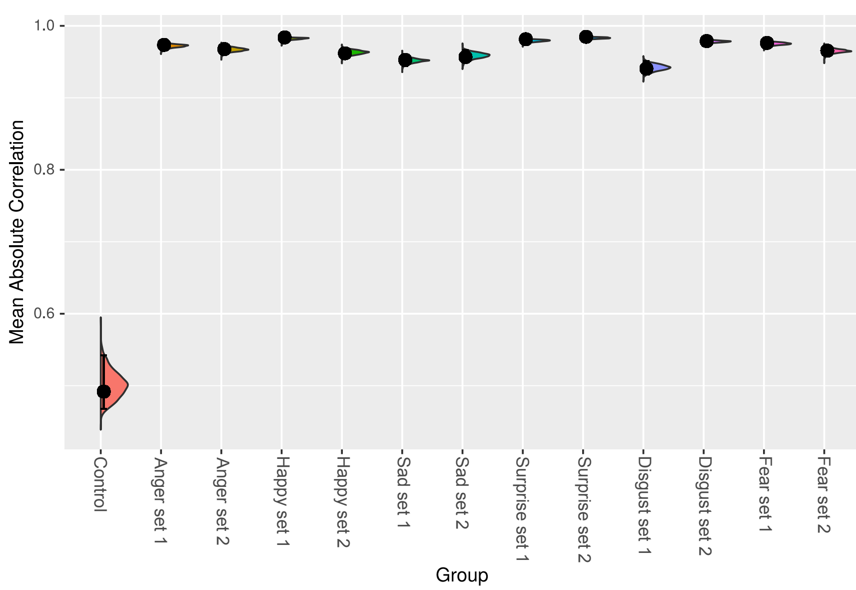


Figure S7: Measures of deviations from perceptual independence, together with their 95% bootstrap confidence intervals. The leftmost plot represents the benchmark of perceptual independence violations provided by the control group, and the rest represent perceptual independence violations observed for all experimental groups.

References

Blunden, A. G., Wang, T., Griffiths, D. W., & Little, D. R. (2015). Logical-rules and the classification of integral dimensions: individual differences in the processing of arbitrary dimensions. *Frontiers in Psychology*, 5, 1531.

Cornes, K., Donnelly, N., Godwin, H., & Wenger, M. J. (2011). Perceptual and decisional factors influencing the discrimination of inversion in the Thatcher illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 645.

Ekman, P. & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues.* Unmasking the face: A guide to recognizing emotions from facial clues. Oxford, England: Prentice-Hall.

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). Not all morph spaces stretch alike: How category learning affects object discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 807–802.

Goldstone, R. L. & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, 130(1), 116.

Lesmes, L. A., Lu, Z.-L., Baek, J., Tran, N., Dosher, B. A., & Albright, T. D. (2015). Developing Bayesian adaptive methods for estimating sensitivity thresholds (d') in Yes-No and forced-choice tasks. *Frontiers in Psychology*, 6.

Linares, D. & López-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, 8(1).

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, (pp. ISBN 91–630–7164–9).

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 328–342.

Soto, F. A. & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, 139, 105–129.

Soto, F. A., Zheng, E., Fonseca, J., & Ashby, F. G. (2017). Testing separability and independence of perceptual dimensions with general recognition theory: A tutorial and new R package (grtools). *Frontiers in Psychology*, 8.

Townsend, J. T. & Wenger, M. J. (2014). On the dynamic perceptual characteristics of Gestalten: Theory-based methods. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization Get access Arrow* (pp. 948–968). Oxford, UK: Oxford University Press.