Face shape and motion are perceptually separable: Support for a revised model of face processing

Emily Renae Martin, Jason S. Hays, and Fabian A. Soto
Department of Psychology
Florida International University
11200 SW 8th St, Miami, FL, 33199

A recent model of face processing proposes that face shape and motion are processed in parallel brain pathways. Although tested in neuroimaging, the assumptions of this theory remain relatively untested through controlled psychophysical studies until now. Recruiting undergraduate students over the age of 18, we test this hypothesis using tight control of stimulus factors, through computerized three-dimensional face models and calibration of dimensional discriminability, and of decisional factors, through a model-based analysis using general recognition theory (GRT). Theoretical links between neural and perceptual forms of independence within GRT allowed us to derive the *a priori* hypotheses that perceptual separability of shape and motion should hold, while other forms of independence defined within GRT might fail. We found evidence to support both of those predictions.

Keywords: Face perception, General Recognition Theory, face shape, face motion

Information about a number of properties can be extracted from a single face, such as identity, emotional expression, race, gender, and movement of facial features. The idea that independent channels process these different dimensions is theoretically appealing, as they would provide streamlined processing of face dimensions that seem to vary independently in the environment, facilitating selective attention and generalization of knowledge acquired about one dimension across changes in the others. Bruce and Young (1986) proposed that emotional expression and identity are processed through parallel routes in the visual system, and later Haxby and colleagues (Haxby et al., 2000) proposed a related neural model, in which face processing begins in the inferior occipital gyrus, from which two parallel streams branch out. Processing in the ventral stream continues to the lateral fusiform gyrus where the fusiform face area (FFA) is housed and where invariant aspects of faces (e.g., identity) are processed. Processing in the dorsal stream continues to the superior temporal sulcus (STS), where changeable aspects of faces (e.g., expression) are processed (Haxby et al., 2000).

Conflicting results from studies following the Haxby et al. (2000) paper have led to the consensus in the literature that revision is necessary (e.g., Bernstein & Yovel, 2015; Lander & Butcher, 2015). Duchaine and Yovel (2015) proposed a revised version of the Haxby model, in which they retained the concept of the ventral and dorsal streams of face processing but updated what face information is assumed to be processed by each stream. Face shape rather than invariant aspects of faces would be represented in the ventral stream, and face motion rather than changeable aspects of faces would be represented in the dorsal stream. Therefore, any motion information connected to an identity would be processed in the dorsal stream and any shape information connected to an emotional expression would be processed in the ventral stream.

Importantly, Duchaine and Yovel (2015) based their hypothesis largely on the results of lesion and neuroimaging studies. The psychophysical literature at present suggests that motion and shape are not behaviorally independent, based on the finding that recognition of both face identity and expression is often facilitated by motion information (Krumhuber et al., 2013; Lander & Butcher, 2015). It is tempting then to conclude that the existence of parallel routes at the neural level does not have corresponding behavioral consequences. However, facilitation effects can result from decisional rather than perceptual processing, as long as participants adopt optimal decision strategies (Maddox & Ashby, 1996). This underscores the importance of controlling for decisional strategies when studying perceptual interaction in psychophysical studies. Here, we test the hypothesis of independent processing of shape and motion using controlled psychophysical experiments and model-based data analyses that dissociate perceptual from decisional forms of independence. Thus, our goal was not only to test this revised model using behavioral psychophysics, but also to use precise, formal definitions of independence, control for nonperceptual influences on task performance (i.e., decisional strategies), and tightly control face stimuli parameters.

Regarding precise, formal definitions of independence, we chose to use the general recognition theory (GRT) frame-

work (for a review, see Ashby & Soto, 2015), a multidimensional extension of signal detection theory (SDT; Green & Swets, 1966) that allows for a formal dissociation between perceptual and decisional forms of dimension interaction. The theory proposes three forms of dimensional interaction, with performance in most tasks being influenced by all three. While the notions of perceptual separability and perceptual independence are perceptual in nature, the third form of interaction, decisional separability, is decisional in nature. To precisely control motion and shape information within the stimuli, we used realistic three-dimensional computer face models (Hays et al., 2020) to create the stimuli.

Defining shape as static changes in inner face features and motion as dynamic changes in the speed of face feature movement (*Figure* 1), we conducted a highly controlled experiment assessing violations of formally defined forms of perceptual interaction.

A Priori Hypotheses

The revised independence hypothesis suggests that separate neural populations encode face motion and shape. Under such conditions, we would expect from neurocomputational theory (Soto et al., 2018) that perceptual separability of the dimensions would hold, which is our first hypothesis. When perceptual separability holds, the representation of one property (e.g., shape) is stable across changes in a second property (e.g., motion). Thus, perceptual separability formalizes the concept of invariance, which is widely believed to be a computational goal of the visual system (e.g., Rust & Stocker, 2010).

On the other hand, there is evidence showing that areas in the two pathways are not completely independent but interact with one another. Assuming that such interactions would introduce correlated neural noise, we hypothesized that perceptual independence would fail. Perceptual independence fails when perceptual noise in one dimension (e.g., shape) is correlated with perceptual noise in the other dimension (e.g., motion). Decisional separability was not expected to hold since decisional strategies may vary by individual. These hypotheses were pre-registered before data collection (see: `https://osf.io/ud7jg?view_only=e6279964d33244b6a6692f02334ae0f0`).

Method

Participants

Three-hundred and ninety participants over the age of 18 were recruited from Florida International University using Sona Systems. Participation was entirely voluntary, and participants were compensated in course credit. All procedures were approved by the university's institutional review board. One-hundred and eighty participants completed the preliminary pilot task (detailed below), whereas the remaining two-hundred and ten participants completed the main identification task.

Sample Size

Model-based analyses like those reported here are not amenable to traditional power analyses to determine sample size. A more common approach in the computational modeling literature is to perform simulations to determine at what sample sizes it is possible to recover the true model parameters that have generated a data set. Such simulations (Soto et al., 2021) suggest that a sample size of 20-30 participants per experiment is adequate for accurate parameter estimation. Therefore, 30 complete datasets were collected for each group.

Stimuli

Each face stimulus was created by concatenating multiple static images into a one-second video, at sixty frames per second. The images were rendered from three-dimensional face models using the software MakeHuman (`www.makehumancommunity.org`) extended with the package FaReT (Hays et al., 2020; `https://github.com/fsotoc/FaReT`). The sequence of renders was obtained by changing the expression pose parameters of a face model in multiple steps, going from no expression to a target expression. FaReT is an open-source toolbox recently developed to aid in the creation of realistic computerized face stimuli. It allows for standardization of non-facial features and exact interpolation of facial features. Additionally, FaReT includes a collection of empirically validated models of identity and emotional expression (Hays et al., 2020).

Face shape was statically manipulated by changing facial feature parameters (e.g., eye shape, mouth shape, etc.), which changed the model's face identity but allowed for orthogonal manipulation of motion information as described below. One stimulus set was comprised of two different male identities, and the other stimulus set was comprised of two different female identities. Only inner facial features (i.e., related to nose, mouth, eyes, forehead, chin, and cheeks) were allowed to vary across models, with other face dimensions (i.e., head shape, neck, and ears) standardized to match the average face model defined in the FaReT database.

Note that we have operationally defined shape changes as changes in the face shape parameters defined within MakeHuman, which are perceived as changes in face identity and referenced synonymously going forward. Motion was operationally defined as the synchrony of movement in the top-half and bottom half of the face. In other words, the dimension of motion had two levels: synchronous movement, defined as matching speed of movement in the top half and bottom half of the face, and asynchronous movement, defined as non-matching speed of movement in the top half and bottom half of the face.

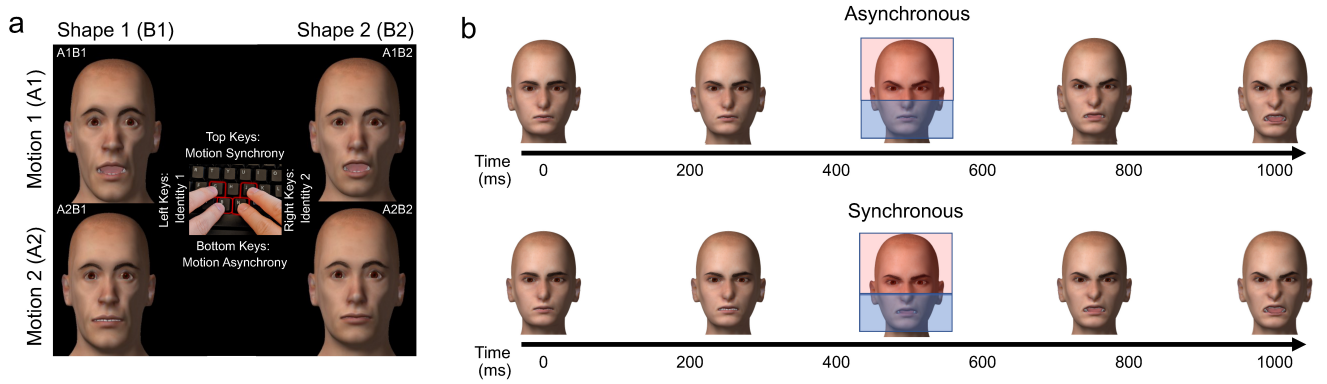

Figure 1. (a) An example of single frames from the dynamic stimuli used for the Male – Surprise group. Participants had to identify whether motion was synchronous (A1) or asynchronous (A2) and whether shape corresponded to one identity (B1) or the other (B2). Note that the synchronous frame shows expression in the entire face and the asynchronous frame only shows expression for the top half of the face. The center sub-panel explains the keys participants selected for each stimulus. (b) Schematic representation of the different motion properties applied to our stimuli; the red box denotes the top half of the face, and the blue box denotes the bottom half. Focusing on the 50% (500ms) point of each speed condition (i.e., the colored boxes), note that the top half has moved equally as much in both conditions (asynchronous and synchronous), but the bottom half has moved only in the synchronous condition, remaining neutral in the asynchronous condition.

There is some evidence that violations of separability are more easily detected in less discriminable stimuli (Wang et al., 2013). To avoid any artificial "asymmetric separability" results, discriminability of the two dimensions should be equated. A pilot study allows for the calculation of dimensional level values that are equally discriminable across each emotion and identity. For the main identification task, both motion and shape were calibrated to have a pre-defined discriminability ($d' = 1.5$; where d' is a measure of perceptual sensitivity from SDT; Green & Swets, 1966), using information from the pilot study. One level of shape was set to one specific identity (e.g., Bob), and the other level was chosen from the Bob-Joe continuum to have a pre-defined discriminability with the first stimulus ($d' = 1.5$). Likewise, one level of motion was set to 0% asynchronous (i.e., 100% synchronous), whereas the other level was chosen from the synchrony continuum to have the same pre-defined discriminability with the first stimulus ($d' = 1.5$). See *Figure 1* for an example.

The faces interpolated from a neutral expression into one of three emotional expressions: happiness, disgust, and surprise. These emotions were chosen from Ekman's six basic expressions (Ekman & Friesen, 1975) and for being the furthest away from each other in the expression space of MakeHuman (in Euclidean distances). Although expression was changed across groups, it was held constant within each group's stimulus set. This allowed us to check the generalizability of results across expressions. To generalize across identities and sex, a second stimulus set of two female identities was included. The shape parameters for these female models were different from the corresponding parameters of the male models (in FaReT, the exact same shape model can be transformed to a male or female), meaning that the stimulus sets differed both in identity/shape and sex.

Pilot

As previously explained, a pilot study was conducted to find the values of stimulus levels that were equally discriminable across categories. An SDT model of the psychometric function (Lesmes et al., 2015) was used to analyze discrimination performance as a function of linear differences between faces in the shape parameter space of FaReT. The estimated function was used to determine the difference between faces producing a $d' = 1.5$ on average. Further information about the pilot study can be found in the Supplementary Materials.

Identification Task

A complete identification task was used, as it can dissociate between all forms of independence defined in GRT (Soto et al., 2015). On each trial of this task, a stimulus with a value on each of two dimensions, A and B, is presented. The task is to identify the specific combination of dimension values presented in the stimulus. For example, consider a 2×2 face identification experiment, where the two varying dimensions are motion (A) and shape (B), as depicted in *Figure 1a*. The levels for the motion dimension are synchronous speed (A1) and asynchronous speed (A2), whereas the levels for the shape dimension are identity one (e.g., Bob; B1) and identity two (e.g., Joe; B2). The 2×2 complete identification task would include four face stimuli: synchronous motion identity one (A1B1), asynchronous motion identity one (A2B1), synchronous motion identity two (A1B2), and asynchronous motion identity two (A2B2).

Procedure

Participants were semi-randomly assigned to one of seven groups, dependent on which of seven links they selected when recruited for participation in the Sona website. They received no information about what specific group they were selecting before beginning the experiment.

Participants completed an anonymous Qualtrics (`https://www.qualtrics.com`) survey to provide informed consent and answer basic demographic questions. Upon completing the survey, participants were redirected to the study hosted on Pavlovia (`https://www.pavlovia.org`), an online server that allows for remote secure data collection, where they would complete the experiment remotely. They were instructed to use either a laptop or desktop computer.

Previous research has found that unfamiliar faces are processed differently than familiar faces (Burton & Jenkins, 2011). Familiarization with faces through exposure to animated videos has been shown to not only help participants learn the identity quickly, but also prevents image-matching strategies, commonly displayed with unfamiliar faces, from occurring (Burton & Jenkins, 2011). Therefore, the task began with a familiarization phase, which presented an animated video of the faces from different viewpoints together with the assigned name of the identity, a strategy that has been successful in previous research (Megreya and Burton, 2006). Each animation was repeated three times and participants were instructed to memorize the faces. Then, the same animations were presented without any names, and the participant was required to choose the correct name from ten options to ensure that they had learned the identities.

After the familiarization phase was completed, the main task began. On a given trial, participants were shown one of the four stimuli (i.e., A1B1, A1B2, A2B1, or A2B2) defined as one of two identities showing one of two motion sequences (i.e., synchronous or asynchronous) and were instructed to identify the face accordingly (i.e., the specific combination of motion and shape, see the middle insert in *Figure* 1a). The pattern of correct and incorrect responses provides information to perform a model-based analysis with GRT, detailed in the analysis section.

There were 460 trials in the identification task and the stimuli were presented for 1s. A fixation cross appeared for 500ms at the beginning of each trial. Feedback was given following each trial; if the response was correct, a green "Correct!" appeared on the screen for 1s, but in incorrect trials, a red "Incorrect!" appeared, followed by a 5s penalty timeout. If the participant provided no response within a 2s window, a red "Too Slow!" appeared for 1s, along with the 5s penalty timeout.

Control Group

In our stimuli, motion was manipulated by changing only one half of the face (i.e., the bottom). To increase the generalizability of our results, we collected data from an additional control group in which motion was manipulated by changing the top half of the face. This group had all the same experimental parameters as the other identification task groups, except or the manipulation of synchrony by changing motion in the top half of the face. The first stimulus set showing an expression of surprise was arbitrarily chosen to define the dimensions for this group. Corresponding pilot data were collected to set dimensional values at $d' = 1.5$, just as before. This group was added post pre-registration[1], but before the analysis was conducted.

Analysis

GRT is a framework used to study steady-state behavior rather than learning. For this reason, it is imperative to eliminate data from each individual that contain evidence of learning, defined by changes in accuracy over time. Therefore, learning curves were obtained for each participant by averaging the performance over a moving window of 101 trials (e.g., trials 1-101, 2-102, etc.). An exponential curve was fitted to the resulting average points using least squares minimization. The learning period was defined as data points before the point where the slope dropped below 0.001, which were discarded from the data of each participant. This method has been implemented and proved useful in multiple previous studies (e.g., Soto et al., 2015; Soto et al., 2021).

A near-perfect performance is not ideal for model-based analyses with GRT, as a lack of confusion errors would not provide any information to fit the model. However, low overall accuracy is also not ideal, as it indicates participants are performing near chance level and did not properly learn the task. Therefore, overall accuracy was calculated after discarding the learning period, and data from participants with performance lower than 40% or higher than 90% correct were excluded from the analysis.

Data from the identification task were summarized in a 4x4 confusion matrix for each participant, in which each row corresponds to one stimulus and each column corresponds to one possible response. Each cell in this matrix represents the frequency of each response after presentation of each stimulus.

Confusion matrices were analyzed by fitting a GRT-wIND model (GRT with individual differences; Soto et al., 2015) using maximum likelihood estimation with the R package *grtools* (Soto et al., 2017). When fitting the GRT-wIND model, the optimization algorithm was run 120 times with random starting parameter values, and the model with the highest

[1] We thank Dr. Lorraine Bahrick for suggesting the addition of this group.

maximum likelihood was chosen as the best-fit model. All variances in the model were fixed to a value of one. Finally, likelihood ratio tests were performed as implemented in *grtools* to assess for violations of perceptual separability, perceptual independence, and decisional separability. The null was defined as the model without these violations, so that a significant result meant that there was evidence for a violation (i.e., an interaction between dimensions).

Results

The first thirty participants that completed the entire experiment and had an accuracy rate between 40-90% (guessing rate = 25%) after the learning period were included in the study. *Figure 2* contains information on how to interpret GRT-wIND outputs for a 2x2 identification experiment. As in SDT, the representation of a stimulus in GRT is probabilistic across trials (i.e., the same stimulus produces different magnitudes of sensory evidence across trials) and its distribution can be summarized by a contour (the ellipses in *Figure* 2). Each axis represents a single dimension (e.g., shape), thus each contour is a multidimensional representation of the perceptual distribution for a single stimulus (e.g., asynchronous Bob). The plus sign inside the contour represents the mean of the distribution. Marginal distributions (i.e., univariate Gaussian curves along the axes) represent the distribution of sensory evidence along a single dimension. For completeness, nonparametric GRT tests of marginal response invariance and sampling independence are reported in the Supplementary Material (see Supplementary Table 2).

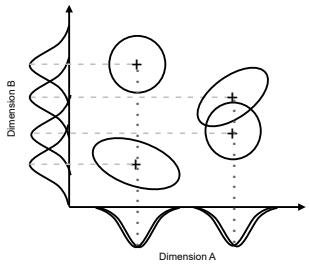

Figure 2. A multivariate Gaussian GRT model for a 2x2 identification task. Each ellipse represents contours of equal likelihood for a specific stimulus's perceptual distribution. The univariate Gaussian curves represent marginal distributions for each stimulus. The darker dotted lines represent the means of the marginal distributions for dimension A, whereas the lighter dashed lines represent the means of the marginal distributions for dimension B. Both of them intersect at the mean of the multidimensional joint distribution (crosses inside each contour). In this example, perceptual separability of dimension A from dimension B holds. That is, the means along dimension A are the same across levels of dimension B. On the other hand, perceptual separability of dimension B from dimension A fails. That is, the means along dimension B are shifted at different levels of dimension A. The two tilted ellipses demonstrate examples of failure of perceptual independence (i.e., positive and negative correlations).

Table 1 shows a summary of the results of likelihood ratio tests evaluating perceptual separability, perceptual independence, and decisional separability. Ultimately, we found perceptual separability of shape from motion in all groups, and perceptual separability of motion from shape in five out of seven groups. These results support our main hypothesis of perceptual separability for the most part. In addition, there was evidence against perceptual independence in all seven groups, and six out of seven groups found at least one violation of decisional separability, thereby showing support for our secondary hypotheses as well. *Figure 3* shows the best-fitting GRT models obtained for each experimental group. The bottom-left insert within each plot depicts a comparison of the performance observed (x-axis) and predicted by the model (y-axis), measured as response probabilities. The main plot contains the perceptual distributions of the model, with each contour corresponding to one of the presented stimuli (i.e., red: synchronous identity 1, green: asynchronous identity 1, blue: synchronous identity 2, yellow: asynchronous identity 2). The corresponding marginal distributions for each stimulus are represented by univariate Gaussian distributions along the corresponding axis. The following sections provide a detailed description of the model fit and likelihood ratio tests summarized in *Table* 1 and *Figure* 3.

Happy Female Set

The model accounted for 98.84% of the variance (R^2) in the data. There was a violation of perceptual separability of motion ($\chi^2(2)=6.33$, $p=0.042$), but not shape. Additionally, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=34.51$, $p<0.001$), which is depicted as tilted contours in *Figure* 3. There was also a significant violation of decisional separability of motion ($\chi^2(30)=57.83$, $p=0.002$), but not shape.

Happy Male Set

The model accounted for 99.12% of the variance (R^2) in the data. The likelihood ratio tests did not find significant violations of perceptual separability for either shape or motion. However, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=41.15$, $p<0.001$), which is depicted as tilted contours in *Figure* 3. Additionally, there was a significant violation of decisional separability of shape ($\chi^2(30)=60.07$, $p=0.001$), but not motion.

Disgust Female Set

The model accounted for 98.76% of the variance (R^2) in the data. The likelihood ratio tests did not find any significant violations of perceptual separability for either shape or motion. However, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=11.11$, $p=0.025$), which is depicted as tilted contours in *Figure* 3.

Group	PS (Motion)	PS (Shape)	PI	DS (Motion)	DS (Shape)
Female Happy	×	✓	×	×	✓
Male Happy	✓	✓	×	✓	×
Female Disgust	✓	✓	×	✓	✓
Male Disgust	×	✓	×	×	×
Female Surprise	✓	✓	×	×	×
Male Surprise	✓	✓	×	×	×
Control	✓	✓	×	×	×

Table 1

Summary of results of likelihood ratio tests. Each row represents a different experimental group, exposed to a unique combination of one of three emotions (happy, disgust, surprise) and one of two identity sets (male and female). PS denotes perceptual separability, PI denotes perceptual independence, and DS denotes decisional separability. Each check mark represents a lack of dimensional interaction (PS, PI, or DS) found for a given group and each cross represents a significant dimensional interaction (i.e., violation of PS, PI, or DS) found for that group.

Additionally, there were no significant violations of decisional separability of either shape or motion.

Disgust Male Set

The model accounted for 98.77% of the variance (R^2) in the data. There was a violation of perceptual separability of motion ($\chi^2(2)=8.69$, $p=0.013$), but not shape. Additionally, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=22.30$, $p<0.001$), which is depicted as tilted contours in *Figure* 3. There were also significant violations of decisional separability for both motion ($\chi^2(30)=67.75$, $p<0.001$), and shape ($\chi^2(30)=118.18$, $p<0.001$).

Surprise Female Set

The model accounted for 98.85% of the variance (R^2) in the data. The likelihood ratio tests did not find any significant violations of perceptual separability for either shape or motion. However, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=35.47$, $p<0.001$), which is depicted as tilted contours in *Figure* 3. Additionally, there were significant violations of decisional separability of both motion ($\chi^2(30)=44.91$, $p=0.039$), and shape ($\chi^2(30)=71.97$, $p<0.001$).

Surprise Male Set

The model accounted for 99.19% of the variance (R^2) in the data. The likelihood ratio tests did not find any significant violations of perceptual separability for either shape or motion. However, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=55.20$, $p<0.001$), which is depicted as tilted contours in *Figure* 3. There were also significant violations of decisional separability for both motion ($\chi^2(30) = 52.27$, $p=0.007$), and shape ($\chi^2(30)=121.45$, $p<0.001$).

Control Group (Surprise Male)

The model accounted for 98.78% of the variance (R^2) in the data. The likelihood ratio tests did not find any significant violations of perceptual separability for either shape or motion. However, the likelihood ratio test for perceptual independence showed a significant violation ($\chi^2(4)=84.44$, $p<0.001$), which is depicted as tilted contours in *Figure* 3. There were also significant violations of decisional separability for both motion ($\chi^2(30)=56.13$, $p=0.003$), and shape ($\chi^2(30)=53.50$, $p=0.005$).

Thus, we found identical results in the control group and its corresponding main group (i.e., surprise male set); both groups demonstrated perceptual separability and violations of perceptual independence and decisional separability.

Akaike Information Criterion (AIC) Weights

The previous analyses showed that perceptual separability held in most cases, but making strong conclusions from null results is questionable. We calculated post-hoc AIC weights (*Table* 2; Wagenmakers & Farrell, 2004) for each of the fitted models to determine the extent to which the data supports a model restricted to show perceptual separability over a model without such restrictions (i.e., the best-fitting model shown in *Figure* 3). We found that the AIC weights for groups not showing any violations of perceptual separability were all above 0.98, indicating that there is very high likelihood that the model with perceptual separability was the best model for those groups. The happy female group with one violation of perceptual separability yielded an AIC weight of 0.909, suggesting that in this case as well there is a very high likelihood that the model with perceptual separability was the best model. The male disgust group with one violation of perceptual separability yielded an AIC weight of 0.755, suggesting that the model with perceptual separability was considerably more likely to be the best model than a model without perceptual separability (precisely 3.08 times more likely). These results show support for the model with

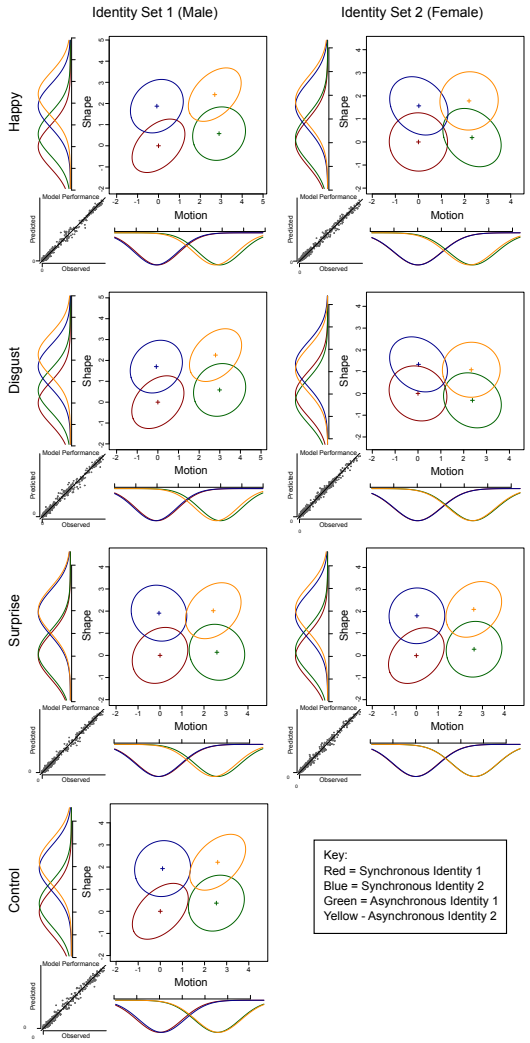

Figure 3. Summary of the best-fitting GRT models obtained from the data of each group in this study. Each panel represents the results for a different group, organized by the emotion (in rows) and the identities (in columns) shown in the stimulus set. The joint perceptual distribution for each stimulus is represented by a contour of a different color (see Key). The plus sign represents the mean of the distribution. The marginal distributions for each dimension are plotted on the corresponding axes. The bottom-left insert plots the performance observed (x-axis) versus predicted by the model (y-axis), measured as response probabilities.

perceptual separability, even in the few cases where we found violations according to the likelihood ratio test.

Discussion

Our study behaviorally tested the revised neurobiological hypothesis (Duchaine & Yovel, 2015) that face motion and shape are processed separately, using tight stimulus control through computerized three-dimensional face models and calibration of dimensional discriminability, as well as dissociation of decisional from perceptual factors through a model-based analysis using GRT. Our results show strong evidence supporting the idea that the parallel neural routes proposed by the Duchaine and Yovel (2015) model have corresponding consequences at the perceptual/behavioral level.

Group	PS of Motion	PS of Shape
Female Happy	*0.9090	0.9971
Male Happy	0.9898	0.9999
Female Disgust	0.9999	0.9999
Male Disgust	*0.7548	0.9999
Female Surprise	0.9990	0.9995
Male Surprise	0.9875	0.9998
Control	0.9968	0.9922

Table 2

*AIC weights for the comparison between models assuming and not assuming perceptual separability. Each row represents a different experimental group, and each column denotes the AIC weight for the model assuming perceptual separability, which can be interpreted as the probability that perceptual separability is closer to the true model than violations of perceptual separability. One minus the presented AIC weight would represent the probability of violations of perceptual separability being closer to the true model. The * symbol represents a significant violation of PS in the likelihood ratio test.*

Although our data suggests that perceptual separability holds, we did find a repeated pattern of perceptual independence violations across the groups. Perceptual independence violations mean that the perceived shape and motion values of a face are correlated. Sometimes, these correlations may show a pattern reflective of holistic or "Gestalt" perception of the two dimensions (Townsend & Wenger, 2014), in which strong perceptual evidence for the presence of a feature (e.g., eye separation) is accompanied by strong perceptual evidence for the presence of the other feature (e.g., asynchronous motion). Using controlled methods similar to those used in the present study, a recent study has found that face identity and expression consistently show that pattern of Gestalt perception (Hosseini & Soto, 2024). Upon closer inspection, we found significant positive correlations in the off-diagonal (i.e., top-right and bottom-left) joint distributions (95% CIs: [0.065, 0.400]; [0.132, 0.408]). This result seems to indicate that one identity may have been easier to perceive with asynchronous motion and the other may have been easier to perceive with synchronous motion, suggesting that different types of motion draw attention to certain facial identity features. However, we did find that most of the correlations were quite weak in magnitude and no other patterns were discernible.

These violations are different from violations of perceptual separability, which occur when changes in the external value of one stimulus dimension influences the internal representation of another dimension. Rather than Gestalt perception, such violations suggest perception of one dimension that is not invariant (i.e., it is context-specific) to changes in

the second dimension. The results are in line with our original interpretation of the neurobiological hypothesis: two independent pathways should produce something akin to invariant representations, but connectivity across the pathways might introduce some noise correlation. The lack of a consistent pattern of correlations across groups suggests that they do not serve a particular perceptual goal (e.g., perceptual integration or holism), and might be the outcome of connectivity across pathways that serves other purposes.

There were also inconsistent patterns of results for decisional separability both within and across groups. This is to be expected from participants showing idiosyncratic decisional strategies. Our main interest is not to interpret these results, but rather to control for decisional factors to draw conclusions about perceptual representation.

We opted for strong stimulus control and chose to create three-dimensional computerized face models, which although are realistic, do not completely match real-world photographs or videos. For instance, our stimuli had a standardized skin texture, despite the fact that texture differences are important for everyday face recognition (Lai et al., 2013). Furthermore, our stimuli were standardized completely across non-facial features, removing any variability not related to the inner shape of the face (e.g., head shape, ears). Standardizing our stimuli and tightly controlling the face parameters allowed us to orthogonally manipulate shape and motion while eliminating potential stimulus confounds, perhaps at the risk of decreased external validity. We believe that this is a risk worth taking, as tight experimental control is key to rigorously test hypotheses regarding the mechanisms of perception and cognition.

We must note that there are other possible interpretations of the two-streams hypothesis in terms of neural encoding. For example, one could assume that the two streams do not interact at all, a case in which we would expect no violations of either perceptual separability or independence (see Soto et al., 2018). The two-streams hypothesis is not detailed and formalized enough to clarify this point, but we have assumed a "less strict" version in which connections between the two streams produce neural noise correlations between them. We have found that only this "less strict" version of the hypothesis is supported by the data. If the two-streams hypothesis posits that there are no interactions between motion and shape at all, then we find evidence against that claim because perceptual independence failed. However, if it posits that there exists some interaction between the pathways that leaves motion and shape separable across stimuli (i.e., invariant representations), then we find strong support for this hypothesis.
possible interpretation of the current data posits that the separation of neural encoding occurs not in the visual processing streams but in later brain areas (e.g., frontal cortex). In Figure 4, we propose two neural encoding models , in which separation may occur early during visual processing (i.e., Duchaine & Yovel's model), or after visual processing and before the encoded information reaches decision making. Following Townsend et al. (2020; see Figure 8), and in line with the neuroscientific literature, both models propose shared noise during an early stage of processing (e.g., the occipital face area), leading to violations of perceptual independence. Further computational neuroimaging studies (e.g., using encoding modeling or decoding analyses) are necessary to arbitrate between these two hypotheses. The current neuroimaging data supports the early model, but studies using formal definitions of neural independence that can be linked to our results (see Soto et al., 2018) are lacking. We hope that our study shows how a rigorous methodology, based on strong experimental control and precise definition and quantification of interactions through model-based analyses, can help to bridge the gap between the neurobiological and psychophysical literatures.

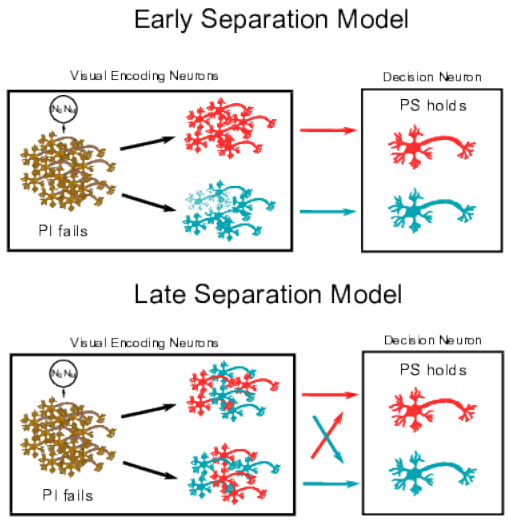

Figure 4. Simplified neural encoding models that our data could explain, in which both perceptual independence fails and perceptual separability holds. In each model, a population of vision neurons that encode both shape and motion in faces exist in early visual processing. Here, perceptual independence fails (i.e., correlated noise within dimensions). In the early separation model, during visual processing, a population of neurons that encode shape (i.e., red neurons) is independent from the population of neurons that encode motion information (i.e., blue neurons). This information is eventually passed to the associative cortex where a decision is made, and perceptual separability holds. In the late separation model, during visual processing, two populations that encode shape and motion information exist, but the populations each contain neurons that encode shape and motion. Therefore, information from both "streams" are passed to the associative cortex where the decision is made, but the information is separated before it reaches the decision neuron and thus perceptual separability holds. Further computational neuroimaging data is required to test these models, but either model is possible from the current behavioral data.

Open Practices Statement

All data, code, and materials are publicly accessible at this link: `https://osf.io/ur3wm/files/osfstorage?view_only=e6279964d33244b6a6692f02334ae0f0`. The preregistration can be found here: `https://osf.io/ud7jg?view_only=e6279964d33244b6a6692f02334ae0f0`.

References

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

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

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

Burton, A.M., & Jenkins, R. (2011). Unfamiliar face perception. In *The Oxford Handbook of Face Perception,* 28, 287–306.

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

Ekman, P., & Friesen, W.V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues* (Vol. 10). Ishk.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

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

Hays, J., 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, 2604–2622.

Hosseini, S., & Soto, F. (2024). Multidimensional Signal Detection Modeling Reveals Gestalt-Like Perceptual Integration of Face Emotion and Identity Emotion. Manuscript accepted for publication in *Emotion*.

Krumhuber, E.G., Kappas, A., & Manstead, A.S.R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5*(1), 41–46.

Lai, M., Oruç, I., & Barton, J. J. (2013). The role of skin texture and facial shape in representations of age and identity. *Cortex, 49*(1), 252-265.

Lander, K., & Butcher, N. (2015). Independence of face identity and expression processing: exploring the role of motion. *Frontiers in Psychology, 6*(255), 1–6.

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, 1070.

Linares, D., & Lopez-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, 8, 122–131.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification–speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 795.

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

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current Opinion in Neurobiology, 20*(3), 382--388.

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., Stewart, R.A., Hosseini, S., Hays, J., & Beevers, C.G. (2021). A computational account of the mechanisms underlying face perception biases in depression. *Journal of Abnormal Psychology, 130*(5), 443–454.

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., 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, 88–111.

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 Perception Science, 8*(696), 1–18.

Townsend, J., & Wenger, M. (2014). On the dynamic perceptual characteristics of gestalten: Theory-based methods. In J. Wagemans (Ed.) *The Oxford Handbook of Perceptual Organization*. Townsend, J. T., Liu, Y., Zhang, R., & Wenger, M. J. (2020). Interactive parallel models: No Virginia, violation of miller's race inequality does not imply coactivation and yes Virginia, context invariance is testable. *The Quantitative Methods for Psychology, 16*(2), 192-212. Ungerleider, L.G., & Haxby, J.V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4, 157–165.

Wagenmakers, E., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196.

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 Emotion Science, 4*(943), 1–11.

Supplementary Materials for "Face shape and motion are perceptually separable: Support for a revised model of face processing"

Contents

1 Definitions of Independence in GRT

Implementing a GRT model (Ashby & Townsend, 1986) to describe psychophysical data gives the ability to dissociate perceptual from decisional processes in the processing of stimulus dimensions. There are three primary concepts describing interactions between dimensions in GRT: perceptual independence, perceptual separability, and decisional separability (see Figure S1).

Perceptual independence refers to the case in which sensory evidence for a stimulus' level in one dimension is uncorrelated with the sensory evidence for the level in a second dimension. For example, in Figure *S1A*, the left figure shows a stimulus with perceptual independence; the representation of identity is uncorrelated with the representation of motion. The right figure shows a stimulus with violation of perceptual independence; the internal representation of identity is correlated with the internal representation of motion.

Perceptual separability refers to the case in which the perceptual representations along one stimulus dimension do not change with changes in the other dimension. In the left panel of Figure S1*B*, the perceptual distributions align along the y-axis and the marginal distributions overlap along the x-axis, showing perceptual separability of motion. Perceptions of motion are not influenced by changes in identity. The right figure shows a violation of perceptual separability; that is, perceptions of motion are influenced by changes in identity.

Finally, decisional separability refers to the case in which the decision bound used to classify values (e.g., level of synchrony) in one dimension (e.g., motion) does not change with changes in the other dimension (e.g., shape). To clarify, a decision bound can be conceptualized as a line that defines at which point the participant begins to choose one level of the dimension instead of the other level. In both panels of Figure *S1C*, this bound is represented by the blue line. In the left figure, the bound is orthogonal to the motion axis, indicating decisional separability of motion. This means that decisions about motion do not depend on identity. The right figure is an example of a violation of decisional separability; that is, an individual

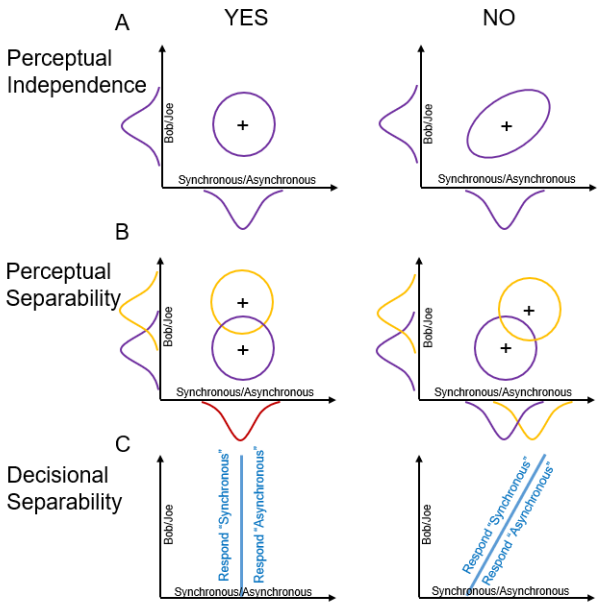

Figure S1: Different forms of dimensional interaction defined under GRT (adapted from Soto et al., 2017).

is more likely to respond "synchronous" when Joe's face is shown and "asynchronous" when Bob's face is shown, demonstrating a bias that depends on identity.

Based on a neurocomputational version of General Recognition Theory (Soto et al., 2018), we hypothesized that perceptual separability would be the most likely consequence of separate brain pathways for the processing of face shape and motion (for a more detailed explanation, see the main text).

2 GRT-wIND Model

The GRT-wIND model (Soto et al., 2015) controls decisional factors by identifying and estimating separate parameters for perceptual and decisional mechanisms. The model takes a hierarchical form, in which group-level parameters explain processes assumed to be shared across participants and individual-level parameters explain individual differences.

Within our 2x2 identification task, stimuli could take on one of two levels of shape and one of two levels of motion. We can capture the perceptual representation through a bivariate Gaussian distribution, described by the mean vector μ_s:

$$\mu_s = \begin{matrix} \mu_{s,x} \\ \mu_{s,y} \end{matrix} \tag{1}$$

and covariance matrix Σ_s:

$$\Sigma_s = \begin{matrix} \sigma^2_{s,x} & \rho_s \sigma_{s,x} \sigma_{s,y} \\ \rho_s \sigma_{s,x} \sigma_{s,y} & \sigma^2_{s,y} \end{matrix} \tag{2}$$

where each stimulus is represented by s, the dimensions shape and motion are represented by x and y respectively, σ is a standard deviation parameter, and ρ is a correlation parameter. All σ were set to 1 to fix the model's scale and avoid identifiability problems that arise in signal detection theory models (Silbert & Thomas, 2017). The parameter μ_{A1B1} was set to [0 0] to fix the position of the model in the two-dimensional space.

GRT-wIND captures individual differences in performance through individual parameters. Parameter κ_k represents the overall attention level shown by participant k, with higher levels decreasing perceptual noise and increasing discriminability along both dimensions. Parameter λ_k represents the selective attention to each dimension shown by participant k, with a value of 0.5 representing equal attention to both dimensions. A value higher than 0.5 represents more attention towards shape and a value lower than 0.5 represents more attention towards motion. Therefore, the covariance matrix becomes modified to:

$$\Sigma_{sk} = \begin{matrix} \frac{\sigma_{s,x}^2}{\kappa_k \lambda_k} & \frac{\rho_s \sigma_{s,x} \sigma_{s,y}}{\sqrt{\kappa^2 \lambda_k (1-\lambda_k)}} \\ \frac{\rho_s \sigma_{s,x} \sigma_{s,y}}{\sqrt{\kappa^2 \lambda_k (1-\lambda_k)}} & \frac{\sigma_{s,y}^2}{\kappa_k (1-\lambda_k)} \end{matrix} \tag{3}$$

Two other individual parameters describe decision bounds assumed to be used by participants. Each bound can be written as a linear discriminant function in which one parameter is always fixed to one and the others define its slope and position:

$$h_{dk}(X, Y) = b_{dk,x} X + b_{dk,y} Y + c_{dk} \tag{4}$$

where d represents dimension (i.e., shape or motion), and X and Y represent specific values of the dimensions x and y respectively.

The two response bounds together subdivide the perceptual distribution associated with stimulus s into four areas that add up to 1.0, each representing the probability of each response (i.e., synchronous ID_1, synchronous ID_2, asynchronous ID_1, asynchronous ID_2) given the stimulus s, $P(R_j|s)$. We can use such probabilities to compute the likelihood of a set of parameters given the data (Soto et al., 2015). We then can estimate the model parameters by maximizing the log-likelihood of the model:

$$LogLik(\theta) = \Sigma_k \Sigma_s \Sigma_j r_{ksj} log P_k(R_j|s, \theta) \tag{5}$$

where r_{ksj} is the frequency in which participant k gave the response R_j when presented with stimulus s and $P_k(R_j|s)$ is the probability of this event given the model parameters, represented by θ.

3 Pilot Methods

Considering the evidence that less discriminable stimuli produce stronger violations of separability (Wang et al., 2013), probably due to selective attention to the more easily-discriminable dimension, we decided to equate the discriminability of the two dimensions. A pilot study was implemented to obtain psychometric functions for shape and motion, which would in turn allow us to calculate the stimulus values leading to a pre-defined sensitivity ($d' = 1.5$) in both dimensions.

3.1 Stimuli

To generate a sequence of face models transforming from one face identity to another in the *shape/identity* groups, models were obtained from the line segment connecting the two identities in the shape parameter space defined within MakeHuman (`www.makehumancommunity.org`) extended with the package FaReT (Hays et al., 2020; `https://github.com/fsotoc/FaReT`). This line segment was divided into equally spaced parts that represent percentages of transformation from one identity to the other, with each percentage point containing its own parameter values for each of the face features manipulated. The collection of models at each step, rendered from a frontal view, became the stimuli. For example, the stimuli for set one consisted of Bob, Joe, and nine images in-between as a shape continuum between Bob and Joe, each in 10% transformation steps. Therefore, 10% Joe (or 90% Bob) translates to the identity being at 10% of the linear trajectory from Bob to Joe. The 50% image became the average face of the two identities, which was used as the constant identity/shape when the target dimension was motion. In sum, we implemented linear interpolation in the parameter space of MakeHuman to obtain the shape dimension.

In the *motion* groups, the parameters of the face expressions were divided into two feature groups: face features around the eyes, eyebrows, and forehead (red region in Figure S2), and face features around the mouth, cheeks, and chin (blue region in Figure S2). For this experiment, we manipulated the level of

synchrony between top face movement (e.g., eyebrow, eyelid) and bottom face movement (e.g., mouth) in 10% steps from complete synchrony (both feature groups moving at the same speed) to complete asynchrony (one feature group moving first completely, followed by movement in the other feature group).

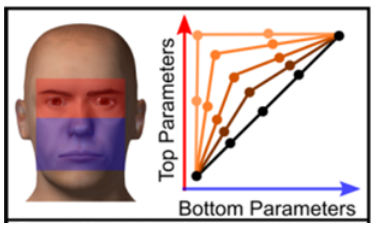

Figure S2: Manipulation of motion asynchrony of face features in dynamic stimuli. The black segment represents complete synchrony, whereas the lightest segment represents complete asynchrony.

The motion dimension was defined as the amount of motion synchrony between the top half and bottom half of the face, just as in the main experiment. As before, we interpolated between complete synchrony (black line in Figure S2) and complete asynchrony (light orange line in Figure S2), to obtain a continuum composed of 11 levels of motion synchrony. The final dynamic stimuli were obtained by concatenating renders of faces transformed from a neutral expression to one of three emotions (i.e., happiness, surprise, and disgust), using the same procedure described in the main text. Emotional expression was held constant within each study group. Stimulus sets presented to different groups varied in sex and identity (i.e., male and female), and emotion (i.e., happiness, surprise, and disgust). Motion and expression were held constant in the identity groups. The faces in these groups displayed a neutral expression and were static images displayed for 200ms. To hold identity constant within the motion groups, the 50% average face created for the shape pilot group was used. Constant texture and illumination were implemented through FaReT's defaults.

3.2 Procedure

There were 48 training trials and 750 subsequent testing trials for the pilot task. During training trials, only the two stimuli representing the endpoints of the dimension (0% and 100%) were presented. During testing trials, the endpoint stimuli (i.e., 0% and 100% conditions) were presented 150 times, and the stimuli created along the dimension (i.e., 10%–90% conditions) were presented 50 times. Trials were organized in 50 blocks of 15 trials each (3 presentations of each endpoint stimulus, and one presentation of each of the other nine stimuli), and randomized within block.

During each trial, a fixation cross appeared for 500ms before the presentation of a stimulus. Static images were displayed for 200ms and the dynamic stimuli were presented for 1s. Feedback was given at the end of trials with stimuli representing the endpoints of the dimension, both during training and testing. If the response was correct, a green "Correct!" appeared on the screen for 1s, but in incorrect trials, a red "Incorrect!" followed by a 5s penalty timeout would appear. If the participant provided no response within a 2s window, a red "Too Slow!" appeared for 1s, along with the 5s penalty timeout.

Participants were instructed to categorize the stimuli based on the target dimension (e.g., categorizing the face as 'Bob' versus 'Joe', or as 'synchronous motion' versus 'asynchronous motion'). The pilot study was divided into nine experimental groups, each with their own target dimension (i.e., male identity, female identity, male happy motion, male surprise motion, male disgust motion, female happy motion, female surprise motion, female disgust motion, and a second male surprise motion control group).

3.3 Data Analysis

Participants showing a low performance level (<75% correct) in the discrimination between the extreme levels of the dimension (e.g., 0% Bob versus 100% Bob) were excluded from the analysis, as we assumed that they did not learn the task or were not paying attention. Psychometric curves that do not reach values close to floor and ceiling of performance cannot be adequately analyzed through curve-fitting. Model parameters

then become difficult to recover, with the fitting procedure often failing to converge to a reasonable set of parameter estimates. Therefore, it is important to discard participants with low performance level to prevent such issues.

Categorization responses were used to estimate psychometric curves for shape and motion at the individual level. The estimated parameters were averaged to obtain psychometric curves at the group level. The psychometric curves were estimated using a signal detection theory model (Lesmes et al., 2015), which breaks down participant performance in terms of sensitivity, independently from decisional biases. As indicated earlier, the psychometric curves for the participants were obtained to calibrate the stimulus sets in the main study and to ensure that the discriminability of dimensional changes is equivalent for shape and motion (both at $d' = 1.5$). Data from the pilot studies were analyzed using the R package *quickpsy* (Linares & Lopez-Moliner, 2016).

4 Pilot Results

Table *S1* depicts the stimulus values used for each group in the main identification task according to a d' of 1.5.

Group	Value
Identity Set 1 (Male)	53.912
Identity Set 2 (Female)	37.884
Motion Happy Set 1	58.647
Motion Happy Set 2	59.189
Motion Disgust Set 1	63.316
Motion Disgust Set 2	49.914
Motion Surprise Set 1	61.631
Motion Surprise Set 2	55.865
Motion Control (Surprise 1)	51.402

Table S1: Stimulus value at d' = 1.5 for each dataset. Each value represents the percentage of the second stimulus shown for each dimension and group. Set 1 = Male. Set 2 = Female.

These stimulus values were calculated from the average of the fitted psychometric curves shown in Figure S3. The colored curves represent each participant's data, and the black curve represents the group psychometric curve found by averaging parameters across participants.

5 Post-Hoc Tests

5.1 Binomial Test

Overall, our results indicate support for perceptual separability of face shape and motion; however, we did find two significant violations. Because those two violations were found using a frequentist likelihood ratio test with a 5% false discovery rate, we wondered whether two significant tests out of fourteen is higher than what we would expect from chance, assuming that perceptual separability (i.e., the null hypothesis) holds in general for shape and motion. A post-hoc binomial test revealed that two violations of perceptual separability is not significantly higher than expected by chance with a=0.05, p=0.123. Indeed, the two significant results had relatively high p-values and would not be found significant under a Bonferroni-corrected a=0.0035. For completeness, we tested whether obtaining only two significant violations of perceptual separability out of fourteen is itself significantly lower than chance, under the assumption that perceptual separability does *not* hold for shape and motion in the true data-generating model. Testing this hypothesis requires assuming a particular level of power of the statistical test (e.g., under the null of no separability, and with power of 0.80, the expected number of significant violations of separability would be $0.8 \times 14 \approx 11$). For this reason, we tested the hypothesis using a binomial test for a range of values of power. We found that with power as low as 0.37 we would still reject the hypothesis of violations of separability in our data (i.e., $p<0.05$).

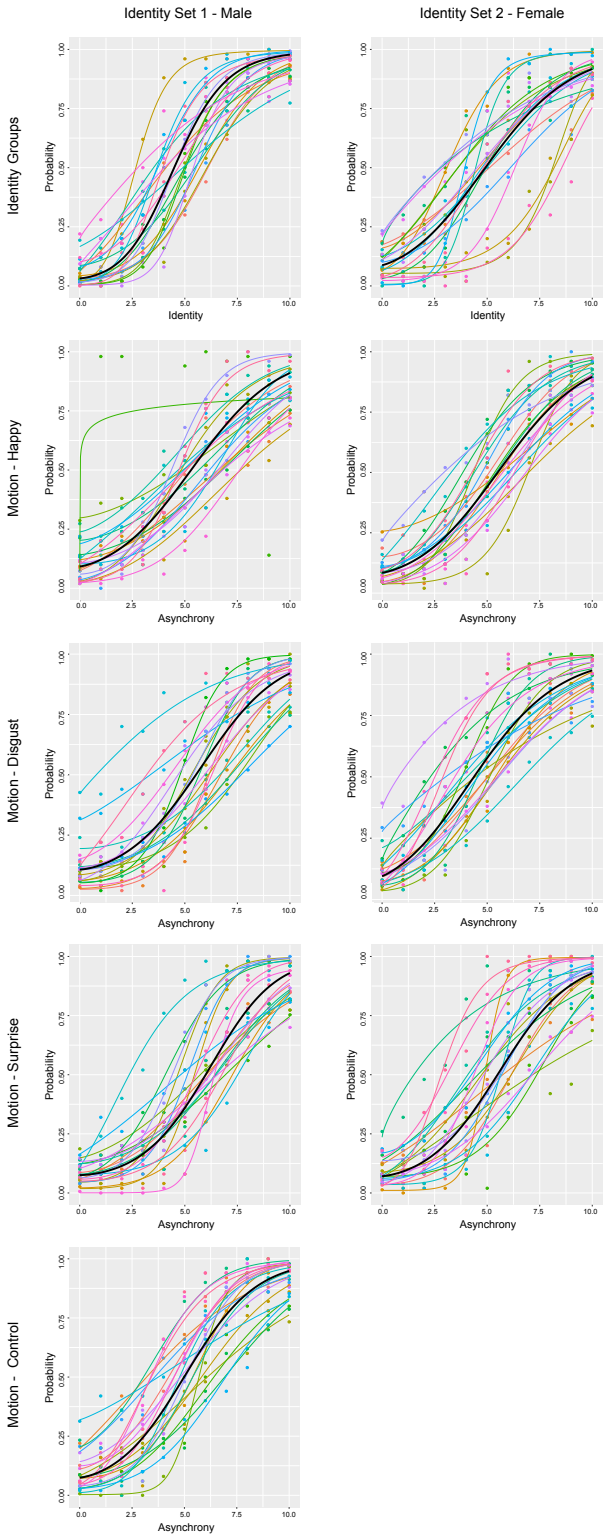

Figure S3: The psychometric curves for each group from the pilot study. Each colored curve represents an individual's curve, and the black curves represent the group average. The curves were fit using a Signal Detection Theory model (Lesmes et al., 2015) with the R package *quickpsy* (Linares & Lopez-Moliner, 2016).

5.2 Nonparametric GRT Tests

Supplementary nonparametric GRT tests of Marginal Response Invariance (MRI) and Sampling Independence (SI) are reported below. MRI is related to perceptual and decisional separability. Because most of our results involved perceptual separability holding but decisional separability failing, we would expect that in most cases MRI should fail. However, we see in Supplementary Table 2 that in the large majority of cases the MRI test holds. Similarly, SI is related to perceptual independence and decisional separability. Because most of our results involved failures of both perceptual independence and decisional separability, we would expect that in most cases SI should fail, but again we see that in the large majority of cases the SI test holds. There are at least two possible explanations for these results. First, a difference in statistical power between our tests with GRT-wIND, which are based on the data from all participants in each experiment, and the nonparametric tests presented below, each based only on the data from a single participant. Violations of decisional separability and perceptual independence that are small in magnitude could very well lead to the pattern of results shown below. As indicated in the main text, this was the case for violations of perceptual independence obtained from our model-based analysis. Second, it has been shown that, in general, analyses of data with GRT focused on individual participants undergoing a 2x2 identification experiment like ours are problematic (Mack et al., 2011; Silbert and Thomas, 2013). GRT-wIND solves these issues when its assumptions are correct (Soto et al., 2015), which is the reason why our pre-registered data analysis applies this methodology.

Group	MRI (Motion)	MRI (Shape)	SI (A_1B_1)	SI (A_2B_1)	SI (A_1B_2)	SI (A_2B_2)
Female Happy	67	77	100	100	100	100
Male Happy	90	80	100	100	100	100
Female Disgust	90	83	100	100	100	100
Male Disgust	80	70	100	100	100	100
Female Surprise	93	53	100	96.7	96.7	100
Male Surprise	93	73	100	96.7	96.7	100
Control	100	70	96.7	100	96.7	100

A = Motion, B = Shape

Table S2: Marginal Response Invariance (MRI) and Sampling Independence (SI) GRT nonparametric tests are reported as the proportion of participants (in percentage points) for which the test held. Therefore, "100" corresponds to MRI or SI holding for 100% of participants in that group.

References

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

Hays, J., 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, 2604–2622.

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, 1070.

Linares, D., & Lopez-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, 8, 122–131.

Mack, M. L., Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Indecision on decisional separability. Psychonomic Bulletin & Review, 18, 1-9.

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

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., 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., 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, 88–111.
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 Perception Science, 8*(696), 1–18.
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 Emotion Science, 4*(943), 1–11.