CATEGORICAL PERCEPTION OF CONSPECIFIC FACES IN RHESUS MACAQUES

BENJAMIN NICHOLAS CASSIDY

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

AUGUST 2013

Abstract

Previous studies indicate humans perceive faces categorically, particularly when the faces are familiar. Categorical perception is traditionally defined by positive results on two psychophysical tasks: an identification and a discrimination task. Whether non-human primates demonstrate the same phenomenon has not yet been explored. This study bridges this gap in the literature by exploring categorical perception of familiar and unfamiliar conspecific faces in two rhesus macaques using computer-generated morph line continua similar to those used in previous face categorization studies. Evidence of both hallmarks of categorical perception was found, demonstrating that rhesus macaques perceive conspecific faces in a categorical manner. This phenomenon becomes more distinct when one, but not both, of the conspecific faces are personally familiar to the individual. Inter-trial adaptation effects cannot account for these results. This study has laid down the behavioural foundation for future exploration of the neural underpinnings of the phenomenon known as categorical perception.

Acknowledgements

I would like to express sincere gratitude to my supervisor, Dr. Kari Hoffman, for sharing her knowledge and expertise over the last three years, and Dr Jen Steeves, for her insight and support throughout the duration of this project.

I would like to acknowledge the contributions and camaraderie of my fellow lab members, past and present, including Adrian Bartlett, Galit Blumenthal, Sonya Chand, Michelle Dragan, Tim Leonard, Xia Liu, Michael Lubinsky, Rodrigo Montefusco, Jonathan Mikkila, Emily Murphy, Roman Skybin, and Josh Tallman.

I would like to acknowledge the team at York University's Animal Care Facility without whose dedication and expertise none of the studies in the lab would be possible, including Natasha Down, Janet Fleites, Julie Panakos, and Veronica Scavo.

I would like to thank all my dear friends, including the 'Neuro-groupies'- Rachel Ganaden, Joe Viviano and Audrey Wong- for their companionship and support, Franca Schawerna for being there at the beginning and the end, and Leah Hartman, who has been a tremendous support and source of sound advice throughout this process.

Lastly, I would like to express love and gratitude to my family, especially Anne and Steve McMillian, for always being a strong foundation and to JP Gloria, for his patience and support throughout this trying period of late nights and long hours.

Table of Contents

List of Tables

List of Figures

The concept of categorization is not a foreign one. All one need do is turn on the television or open a web browser to be bombarded with evidence that humans have a tendency to sort each other into discrete categories based on attributes and characteristics that, in actuality, exist along broad, continuous spectrums. For instance, skin colour varies along a continuum of hues, and yet we often categorize people as 'black' or 'brown' or 'white'. Everyone is placed in a box, from their order at Starbucks and taste in music to their sexual orientation and perceived colour of their skin. These boxes, or categories, are meant to define us as 'individuals', a term riddled with irony given the context. While some embrace the labeling, others resist it; regardless it cannot be denied that categorization plays a central role in how we form impressions of each other and interact socially. That said, envisioning a similar phenomenon on a more basic, perceptual level does not require a stretch of the imagination- just as we may find it easier to define others by the categories in which we place them, in theory, a visual system bombarded with sensory stimuli would do well to organize its percepts categorically. If this were the case, categorical perception could play a pivotal role in the identification of objects, defining how we perceive the world around us, along with the individuals and objects existing within it.

Many millennia ago, certain organisms developed the ability to visually perceive their external world to great evolutionary advantage (Land & Nilsson, 2001). Over time, this visual system has come to serve an important function for many species: the identification and recognition of objects. It parses the visual world into objects and background and then identifies these objects quickly enough to allow the organism to make rapid decisions about the potential threat (or benefit) the objects may pose. However, the underlying neural mechanisms of this system remain somewhat of a mystery. This function necessitates a visual system with a great degree of inherent flexibility and generalization: initial identification of an object must occur within a fraction of a second under varying lighting conditions, angles of observation, distances, degrees of occlusion and other contextual variables, while still operating within the physiological confines of a neural system. In order to make these snap judgments, the visual system could, at least in part, employ some sort of categorization process during visual perception. Physical disparities in an object's appearance caused by the aforementioned contextual variables could be resolved by dropping them into gross categorical bins, instead of taking the time to perceive all the visual nuances of that particular exemplar. For instance, when viewing pictures of animals with the intent of picking out giraffes, it would be inefficient for the system to perceive subtle variations in neck or leg length, fur patterns or colours, beyond what constitutes acceptable values for the category 'giraffe', i.e., all animals with a 'long' neck, 'long' legs, and 'yellowy' fur with 'brownish' spots. Acknowledgment of the minor, physical variations that occur naturally between specimens of this category is not critical for the initial, first-pass identification of the object, unless they are significant enough to cross a category boundary; the neck and legs are short enough and the fur is white enough with dark enough spots to categorize the object as a dairy cow rather than a giraffe.

Liberman and his colleagues (1957) were the first to confirm that humans do indeed perceive some sensory stimuli categorically in their psychophysical studies of speech perception. He noted that speech sounds or 'phonemes', when equally spaced on a physical continuum, are not perceived as varying continuously, but instead are perceptually sorted by participants into specific categories of sounds (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Harris, Hoffman, & Griffith, 1957). Although theories of categorical perception first gained footing in the auditory literature, other scientists soon began to explore whether this phenomenon generalizes to perception in other sensory modalities, specifically vision (Bornstein & Korda, 1984). An excellent demonstration of the categorization of low-level visual stimuli can be seen in the natural world, namely our perception of rainbows (Beale & Keil, 1995).The chromaticity of light varies continuously along the visible spectrum by wavelength (380-750nm; Bornstein, 1987). Although each wavelength differs from the next by a measureable quantitative change and, given the correct psychophysical circumstances, humans are able to discriminate between a good number of them, we identify hues using a relatively small number of hue categories (Bornstein, 1987). As a result, we perceive a rainbow to be compiled of discrete bands of colour, rather than the continuum of hues present (Beale & Keil, 1995). Empirical studies support this anecdotal evidence; colours are indeed perceived categorically (Bornstein & Korda, 1984).

To investigate this phenomenon, researchers in the field established empirical methodology defining what constitutes evidence of categorical perception. The greatest hallmark of categorical perception is a non-linear, step distribution of percepts of stimuli equally spaced along a physical continuum. Within the categorical perception literature, this is referred to as a labeling or identification function. To illustrate this concept, the example of hue perception can be used (Figure 1). In a typical study of hue categorization, a researcher may repeatedly present the participant with colour stimuli of different wavelengths equally-spaced along the visible spectrum between what is prototypically perceived as blue and green (in Figure 1, 500-524nm and 525-550nm, respectively). If these hues were not perceived categorically, it would be expected that the distribution of 'green' responses to hue stimuli would vary linearly across the visible spectrum. In contrast, categorical perception dictates a rather shallow slope in perception of hue followed by a steep 'step' in perception approximately midway along the spectrum, followed again by a shallow slope conveying the idea that 'it's blue until it's green'. This step, or sudden change in perception, is referred to as the "category boundary"; the physical value of stimuli where perception of one category over the other occurs 50% of the time. To generate this distribution, one or both of two psychophysical tasks are often employed: an identification task, in which the participant is asked to identify a stimulus as belonging to one category or the other, or a discrimination task, in which a participant is presented with pairs of stimuli spaced equally along the physical continuum and asked whether they belong to the same or different categories (Cheal & Rutherford, 2010). In the latter task, there are two types of stimulus pairs, those that fall to one side of the category boundary and are thus labeled 'within category' pairs, or those that straddle the category boundary and are thus labeled 'across category' pairs. If the perception of stimuli follow the identification function, within category pairs should be judged to be similar significantly more often than across category pairs, despite the physical differences between the stimulus pairs being equal.

Using this established methodology, vision scientists broadened their investigation beyond low-level visual stimuli, such as hues, to more complex visual stimuli, such as objects, with a notable concentration on the categorization of faces. The idea was that if more complex visual stimuli are also perceived in a similar categorical manner, perhaps categorization is not specific to low-level perceptual processes but rather reflects more general cognitive processes (Beale & Keil, 1995). Faces are a particularly interesting class of objects because humans seem to have a propensity to identify and recognize faces quite rapidly, especially familiar ones (Ramon, Caharel, & Rossion, 2011) from an early age (Diamond & Carey, 1977). From an evolutionary standpoint, faces are extremely biologically relevant stimuli for a social species such as humans, conveying emotion through shifts in musculature (Etcoff & Magee, 1992), or intent through gaze (Perrett & Emery, 1994).

Facial expressions vary along a number of natural continua, such as sad-happy, and thus, in theory, could be perceived categorically. Etcoff and Magee (1992) conducted both discrimination and identification tasks of illustrated faces that varied along multiple physical continua of expression generated by a computer program. It was the first study to employ an, albeit primitive, form of morph lines in the study of categorical perception. A morph line consists of two extremes, or endpoint stimuli, that represent prototypes of the categories being investigated with intermediate morph stimuli generated by a computer algorithm falling on a physical continuum between the two endpoints. In the case of facial expression, all of these intermediate, morph stimuli would occur naturally, but using a morph algorithm allowed researchers to quantitatively control the physical change between stimuli (in this case line drawings representing facial expressions). Similar studies of emotional expression were later conducted using morphs of photographs (Calder, Young, Perrett, Etcoff, & Rowland, 1996; Cheal & Rutherford, 2010; Gelder, Teunisse, & Benson, 1997; Teunisse & Gelder, 2001; Young, Rowland, Calder, Etcoff, Seth, & Perret, 1997). Results from these studies confirmed that facial expression in humans is perceived in a categorical manner.

Initially, categorical perception was thought to only occur for visual stimuli that naturally vary across a physical continuum. Beale and Keil (1995) challenged this idea, investigating whether an individual's face could be treated as a category in and of itself and, if so, if morph stimuli varying across an artificially generated physical continuum between two individual's faces would demonstrate the hallmarks of categorical perception. Computer-generated morph lines of pairs of photographs of famous faces, such as Bill Clinton and John F Kennedy, were presented to participants in both identification and discrimination tasks. The familiarity of the face stimuli was also measured and varied across stimulus pairs. These studies were the first to demonstrate that faces can be perceived categorically and the degree of categorization was relative to the familiarity of the faces being presented (Beale & Keil, 1995). Perhaps one of the most well-known follow-up studies was conducted by Rotshtein, Henson, Treves, Driver, and Dolan (2004), who employed a morph line of Margaret Thatcher and Marilyn Monroe's faces, again choosing highly familiar faces for endpoints. Additionally, Ramon, Caharel and Rossion (2011) recently noted that familiar faces seem to be categorized more quickly than non-familiar faces in a go, no-go task of categorical perception. It would appear that familiarity plays a key role as to whether face stimuli will be perceived categorically, suggesting that face categorization may not be an innate property of the visual system but rather acquired through repeated exposure (Beale & Keil, 1995; but see Campenella, Hanoteau, Seron, Joassin, & Bruyer, 2003).

Although the studies that have been discussed thus far have only explored categorical perception in humans, animal studies have demonstrated similar categorization in a variety of species, including pigeons (Herrnstein & Loveland, 1964), bees (von Frisch, 1964), and non-human primates (Sandell, Gross, & Bornstein, 1979). Animal modeling is an extremely useful methodology, as the neural underpinnings of the human visual system remain relatively unexplored. In contrast, there exists an abundance of research defining the visual system of non-human primates, specifically that of rhesus macaques, however, very few studies have explored the performance of macaques on rapid visual categorization tasks.

In an attempt to bridge this gap in the human and macaque comparative literature, Fabre-Thorpe and colleagues carried out a series of object categorization studies in both macaques and humans over the last decade. The initial study (Fabre-Thorpe, Richard, & Thorpe, 1998), along with the studies that followed, employed a go, no-go rapid categorization task for both species, in which previously unseen natural images of objects belonging (or not) to one of two categories (food or animal) were briefly presented (80ms) on a tactile screen. The participant (macaque or human) was tasked with categorizing the images by either indicating the presence of an exemplar of the category (food or animal) by removing their hand from a button and touching the screen (go response), or keeping their hand on the button to signify the absence of a target object (no-go response). The objects in the images were presented against their natural backgrounds, such as a frog sitting on a leaf in a forest environment, or a piece of fruit on a table, and a wide variety of exemplars were used for both categories (fruits, nuts, and vegetables for food, and a wide variety of species of animals). It is important to note the differences between this categorization paradigm and the traditional methodology previously described here. Fabre-Thorpes and colleagues (1998) did not vary their exemplars of a category equally across a physical continuum in a controlled manner and they provided contextual cues in their images. They also only used one macaque participant per category explored, leaving open the question of generalization across individuals of the species, at least after this initial study.

Remarkably, performance of macaques on the task was good, despite its high demands on the visual system. Upon first exposure to the stimuli, accuracy was slightly lower than humans (90.5% for the food task and 84% for the animal task), while reaction time was slightly faster (356ms for the food task and 251ms for the animal task), with both macaque participants demonstrating a bias for go responses, suggesting perhaps a speed-accuracy tradeoff. Improved performance on repeated trials was negligible, demonstrating little effects of learning and suggesting that a similar process was used by the visual system for both familiar and novel stimuli. It was concluded that macaques must rely on abstract categorical concepts, as with humans, given their comparably rapid categorization of natural images (Fabre-Thorpe et al., 1998).

Further studies were conducted by this group using the same task to rule out the possibility that participants were using low-level cues inherent to the natural images to make their decisions versus abstract categorical concepts. It was established that consistent differences between category and distractor images in colour (Delorme, Richard, & Fabre-Thorpe, 2000), contrast (Mace, Delorme, Richard, & Fabre-Thorpe, 2010), spatial frequency (Girard & Koenig-Robert, 2011) and context (Fize, Cauchoix, & Fabre-Thorpe, 2011), could not account for their results. It would follow that both macaques and humans were categorizing objects (animals and foods) using abstract categorical concepts and were able to generalize to new novel exemplars of the category (Girard, Jouffrais, & Kirchner, 2008). A non-motor, forced-choice saccadic categorization task was also conducted using the same stimuli with similar results (Girard et al., 2008). Overall, these studies have provided considerable evidence that macaques are able to categorize objects in a similar fashion to humans, but neglected to address the idea of controlled physical continua and categorization of intermediate stimuli- all images shown either belonged to the category, or they did not.

Comparatively, little work has been done to explore the idea of face categorization in non-human primates, which is surprising given the overwhelming focus of human literature in this area. One study investigated categorical perception of conspecific and nonconspecific faces in chimpanzees (Martin-Malivel & Okada, 2007). Although evidence has been collected that rhesus macaques process conspecific faces in a similar manner to humans (Dahl, Logothetis, & Hoffman, 2007; but see Parr, 2011), no study to date has attempted to replicate findings in the human literature of face categorization using computer-generated morph lines.

This study bridges this gap in the literature by exploring categorical perception of biologically relevant stimuli, namely conspecific faces, in macaques, using computer-generated morph line continua similar to those used in previous face categorization studies (Beale & Keil, 1995; Cheal & Rutherford, 2010). Preliminary data collected by an Undergraduate Honours Thesis student in our lab suggests that macaques do categorize conspecifc faces in a similar fashion, adhering to the pre-described definitions of categorical perception, at least for familiar conspecific faces. The first hypothesis is that the four rhesus macaques tested in this study will perceive familiar conspecific faces in a categorical fashion, as evidenced by the presence of an identification function and a discrimination effect in their behavioural data. Additionally, given the evidence that familiarity of face stimuli affects categorical perception of these stimuli, it is predicted that the degree of familiarity of the endpoint face stimuli used in this study will bias rhesus macaques categorical perception in such a way that category boundaries will be significantly moved toward nonfamiliar versus familiar endpoint stimuli. This study is intended to lay the behavioural groundwork for future electrophysical studies in the lab, exploring the neural mechanisms underlying the phenomenon of categorical perception of conspecific faces. Studies have shown that clinical populations, such as individuals with autism, demonstrate measureable deficits in these tasks (Teunisse & Gelder, 2001). Results of this study and studies to follow will contribute to our understanding of object recognition in both the intact and lesioned primate visual system.

Method

Subjects

Four adult female rhesus macaques, LE, AN, LU and RI (ages 12, 6, 6, and 6 years, respectively), have been housed socially together for more than five years in York University's non-human primate animal housing facility. Although all four females interact with one another in this set-up, LE has the more visual contact with AN than with LU or RI and vice versa. All procedures conducted with these subjects were approved by the York University Animal Care Committee, which follows the guidelines outlined by the Canadian Council on Animal Care. All four rhesus macaques have served as subjects in previous visual psychophysics experiments, but faces were never differentially reinforced in these experiments.

Task

Stimuli. The categorical face perception task uses 'morph lines' between pairs of photographs of faces. Conspecific faces were used in this study, both familiar and unfamiliar to the subjects, as endpoints along a linear morph continuum. Digital photographs were obtained for each of the four female subjects, along with the male that is in the same colony room with them. The images of unfamiliar conspecific faces were borrowed from a large database of images kindly made accessible by Dr. Katalin Gothard. The background of all images were masked with a uniform mid-grey and the appropriate stimuli were paired. That is, each female familiar face image was matched with another familiar female face and with an unfamiliar female face, creating both a female familiar-familiar pair and a female familiar-unfamiliar pair for each subject's image. Lastly, the familiar male face was paired with an unfamiliar male face. As a result, a total of seven endpoint pairs were created: two female familiar-familiar pairs (RI-LU, AN-LE), four female familiar-unfamiliar pairs, and one male familiar-unfamiliar pair. These pairs were then digitally adjusted using Adobe Photoshop © software, such that face size, image size, contrast and colour of the paired images were similar to each other.

After the endpoint stimuli pairs were established and edited, the images were uploaded in Psychomorph software (Tiddeman, Stirrat, & Perrett, 2005)- a software program that takes endpoint images and creates a continuum of morph images between the two endpoints (see Figure 2). The resulting morph line has proportional elements of each endpoint image changing in a linear fashion over a set number of images (e.g., a morph line with three morph images would produce a 75% 'endpoint 1'/25% 'endpoint two' image, a 50%/50% image, and lastly a 25%/75% image). To guide the morph algorithm, Psychomorph asks the user to create any number of points- and connect these points with lines where appropriate- to form a template overlay of the face image, capturing key facial structures to be used as landmarks in the morphing process. In this experiment, 117 points were placed on predetermined structural features over each of the faces. For example, one point was placed on the centre of each pupil, another on each corner of the mouth, and others around the exterior of the face. After the face templates are created for each face in an endpoint pair, the number of steps, or morph images, is specified and Psychomorph renders the morph images. In this study, nine morph images were rendered for each set of endpoint images (see Figure 2).

Design. Each morph line was presented in a categorical perception task using Presentation software (NeuroBehavioral Systems). The task used in this study was an altered version of the task developed by undergraduate Honours student, Josh Tallman, in a pilot study for this task, which itself was modeled after the visual expectation paradigm used by Cheal & Rutherford (2010). Each trial consisted of the presentation of a stimulus followed by a 2-alternative forced choice scenario. There were two versions of the task, a training version and a testing version. These versions differed from one another only in that the training version presented only endpoint images while the testing version also included morph images. The general structure of a trial is outlined in Figure 3.

Each training trial began with a blank mid-grey screen with centered black crosshairs. After fixating the crosshairs for 250ms, the crosshairs were replaced with one of the two endpoint images of the morph line (spanning 17.23 degrees of visual angle) and a brief tone sounded. The endpoint image was randomly selected with replacement in each trial. The x-axis coordinates of the presented stimulus remained constant from trial to trial (centering the stimulus on the screen), while the vertical coordinates were randomized with replacement from the vertical centre of the screen within a +/- 83 pixel range. This vertical toggle was integrated into the task to prevent the subjects from adopting a strategy involving the use of low-level spatial cues that vary from stimulus to stimulus. After the stimulus was fixated for 75ms, a blank mid-grey screen was presented for 200ms before two checkerboard blocks (each spanning 4.958 degrees of visual angle) appeared 500 pixels to the left and right of the centre of the screen. One of the endpoints 'cued' a left-checkerboard fixation and the other cued fixation of the right checkerboard. When the subject fixated for 200ms on the checkerboard correctly corresponding to the previously presented stimulus, the incorrect checkerboard disappeared and a juice reward was dispensed through the spout, followed by a black screen lasting 1500ms marking the end of the trial. If the subject fixated on the incorrect checkerboard block for the same duration, or neglected to fixate on either of the checkerboards within a 4000ms temporal window, the incorrect checkerboard block disappeared and a buzzer sounded. Lastly, a black screen was presented for 2500ms marking the end of the incorrect trial and forcing the subject to wait a little longer than in correct trials before having the opportunity to earn a juice reward in the next trial. Incorrect trials were always followed by a trial with the same endpoint image, allowing the subject to 'correct' the erroneous choice. The total number of trials presented in any given testing session was flexible and determined by the experimenter.

The testing version of the task was structured in the same manner, with two exceptions: 1) morph images were also presented and 2) the total number of trials presented was fixed. In trials where the presented stimulus was a morph image, the animal received a juice reward regardless of their choice between the left and right checkerboard blocks. These trials served as probe trials and were embedded in a much greater proportion of trials (85%) with endpoint images. Trials with endpoint images had the same structure as those in the training version of the task, that is, feedback was given. Stimulus selection for any given trial in the testing version was structured in such a manner that, for each trial, the stimulus was selected randomly without replacement from a pool of 360 images from a morph line, 15% of which were morph images and 85% of which were endpoint images.

Apparatus. The subjects were transported from their living quarters in the vivarium facility to the lab space using a primate chair, which was placed within a darkened booth in the lab. An LCD monitor (40x32cm using 60Hz refresh with 32-bit colour) was positioned approximately 44cm from the primate chair in the booth. An iViewX infrared eye-tracking system (SensoMotoric Instruments) was also incorporated into the booth with infrared LEDs and a camera centered below the aforementioned screen. The last component to the booth set up is the juice dispenser and spout, which is controlled by programmed commands from Presentation and can also be triggered manually from outside the booth. The spout is fastened to the neckplate of the chair by screws, allowing the subject to lick the juice as it is dispensed.

Behavioural Procedure

Preliminary training paradigm. Subjects' reward preferences were tested by offering multiple types of food/fluid reward and seeing which one was selected first by the animal. LE, AN and LU preferred a juice reward comprised of three parts pear and one part banana baby food, diluted with cherry-flavoured drink, whereas RI preferred pure cherry-flavoured drink. Additionally, a complementary strategy incorporated into this study was a carefully monitored caloric or fluid control program that increased the desirability of the juice reward, thereby increasing the subject's motivation to perform the task. Food and water control adhered to protocol approved by the York University Animal Care Committee. This protocol seeks to minimize the amount of control used, with water control avoided if possible. To summarize briefly, three stages are outlined in this protocol. The first stage uses stimulating images or videos as 'rewards' during experimentation. Given that monkeys habituate to images quite rapidly, this form of reward is often only used intermittently. The second stage replaces treats received in the subjects housing units with treats during testing. When or if this stage does not provide enough motivation, subjects are then moved to caloric restriction in stage three, which is calculated based on base-level caloric intake of that specific monkey and their weight and carefully monitored by a staff veterinarian. Fluid restriction is used as a last resort and is also carefully monitored by vet staff.

The subjects participated in previous behavioural studies in the lab and, therefore, have some basic behavioural training. All four subjects were already trained to be taken out of the colony room and into the lab testing booth, to be calibrated for eye-tracking, and were familiar with the association of receiving a juice reward for various visual behaviours, such as fixating on an object on screen or searching for an object amongst distracters. Both AN and LU participated in an early version of this task in a pilot study conducted six months prior to beginning training for this study. For this reason, AN and LU were the first two subjects selected to be trained on the task, followed by LE and RI after AN and LU data were collected.

The two main challenges to the task that needed to be addressed in the preliminary training paradigm were 1) learning any cue-response mapping and 2) learning the mapping even for complex endpoint stimuli. To address these challenges, subjects were first trained on a version of the task with a pair of simple, distinct geometric shapes (green triangle and orange square) without any decision-making element. One of the two shapes was presented on screen for 75ms, followed by presentation of the correct checkerboard block, which had to be fixated in order to receive the juice reward. Trials were separated into training sets by shape, such that subjects would complete a series of consecutive trials with one shape presented before alternating to a series in which the second shape was presented. In this phase of training, the goal was to instill in the subjects the idea that each of the shape cues was associated with a subsequent fixation location to either the left or right checkerboard block when only one option – at the correct location - was presented on the screen. This continued until the subjects surpassed an accuracy threshold of more than 85% correct trials, at which point choice behaviour was introduced.

In this phase of training, subjects were still only presented with one shape stimulus in any given block of trials, but both checkerboards response squares followed the shape cue, thus the subjects to choose the correct one to fixate to get rewarded. Again, after reaching the accuracy threshold, block lengths were continuously reduced until the subject was able to meet the accuracy threshold while presentation of the two shapes alternated randomly with replacement (as seen in the final version of the task).

After establishing this choice behaviour with the pair of simple geometric shapes, new pairs of stimuli of increasing complexity were introduced in steps, always waiting for the accuracy threshold to be surpassed before moving forward. The 'steps' of complexity of stimulus pairs were as follows: 1) an animal and an inorganic object (a bus and a cow), 2) visually distinct animals (a cheetah and a sheep), 3) animals similar in appearance (a red panda and a fox), and finally 4) unfamiliar conspecific faces. It was determined that subjects had completed preliminary training and were ready to begin working with the test stimuli when performance on multiple different pairs of conspecific faces consistently surpassed the accuracy threshold of 85%.

In addition to gradually increasing the complexity of the task in this manner, there were a few different trial variables that could be manipulated to facilitate training, namely the presence of the buzzer on incorrect trials, the length of the delay (or black screen) after incorrect trials, and juice reward levels. As is the case with all behavioural training, manipulation of these variables was tailored to the individual and varied throughout the training process. For instance, early on in training, limiting the number of trials presented played a key role in maintaining consistent behaviour and motivation, while later on in the training process juice reward levels played a larger role in this process. It is important to note, however, that these variables were only manipulated during preliminary training; all variables were held constant during the training sessions involving morph lines that were later to be tested.

Testing paradigm. Each subject was tested on three different morph lines: a female familiar-familiar morph line, a female familiar-unfamiliar morph line, and a male familiar-unfamiliar morph line (see Figure 4). The identity of the familiar female faces in these morph lines was dependent on the subject's level of visual exposure to the other subjects, balancing for degree of familiarity. The female familiar-unfamiliar morph line featured the familiar female to whom the subject had the most visual exposure, with the other two familiar female faces comprising the female familiar-familiar morph line. As such, LE was presented with the AN-unfamiliar and RI-LU morph lines, LU was presented with the RI-unfamiliar and LE-AN morph lines, and RI was presented with the LU-unfamiliar and AN-LE morph lines. As AN was equally exposed to all three familiar female faces, she was presented with the LE-unfamiliar and LU-RI morph lines to ensure proper balancing of stimuli across subjects. All subjects were presented with the same male familiar-unfamiliar morph line.

For each morph line, one endpoint image (i.e. 100% face identity) is associated with either the left or right checkerboard block; therefore, for morph lines presented to more than one subject, the correct response location was swapped. For example, LE was presented with the RI-LU morph line in which RI cued a left fixation, whereas AN saw the same morph line, but for her, RI cued a right fixation. In the case of the male familiar-unfamiliar morph line, two subjects were presented with a morph line in which the familiar male face was paired with the left checkerboard block, while the other two subjects were presented with a morph line in which the familiar male face was paired with the right checkerboard block.

The testing paradigm is outlined in Table 1. After 10 months of training efforts, only subjects AN and LU reached test level criteria and thus only their data are presented here. Both LU and AN were trained on each morph line using the training version of the task until their performance exceeded the 85% accuracy threshold. For most morph lines, subjects reached this threshold in the first training session over three sets of 120 trials. Afterwards, three testing sessions were completed on separate testing days using the testing version of the task. To ensure the accuracy threshold was being met during testing, each test session was preceded by a short trial set of the training version of the task. Considering each morph line had nine morphs and two endpoints, and that each testing session was comprised of 360 trials of which 85% were endpoint and 15% were morph trials, 162 morph trials (18 presentations/morph image) and 918 endpoint trials were presented for each morph line to each subject over the three testing sessions.

Data analysis

Previous studies investigating categorical perception of faces along morph lines gathered data for a particular morph line from a number of human participants. Each participant's mean response values regarding their perception of the identity of morph images in a morph line served as data points within the sample distribution instead of having to treat their individual categorical responses as sample points themselves- the statistic tested in these studies was the mean of all mean response values across participants and not the mean of their binomial responses. In such a case, the sample distribution is no longer binomial and can be analyzed with traditional parametric (or nonparametric) statistical tests. In the case of these data, each individual's morph line data must be analyzed separately, with their mean response for each morph image on the morph line serving as the statistic to be analyzed. As such, the sample distribution of data points for that mean response are binomial and must be treated as such when analyzed statistically. For this reason, certain limitations are imposed on the analysis of this data. Firstly, only two individuals, LU and AN, were successfully trained on the task and, aside from being shown the same familiar/unfamiliar male face morph line, were otherwise shown unique morph lines. This was necessary given the pool of familiar faces I was able to pull from, but it prohibits pooling data across subjects. Even if data had been collected from all individuals and it was possible to show them all the same set of morph images, the power of the statistical analyses with a sample size of four is limited, even when using nonparametric measures. Given these limitations, I have created analyses that attempt to address the same core concepts of categorical perception, albeit from a slightly different angle. These analyses along with the rationale for using them are described below.

A great deal of data processing occurs between the initial representation of the raw data as text files written by iView software during experimentation and the statistical analysis of variables of interest. All processing and analyses were conducted using MATLAB ™ software. An initial script was written to read through the raw data from iView; identify all fixations, saccades, behavioural data points and descriptors of trial type; and then use the time points to which these various elements were anchored to create a structure in MATLAB that has all these data points sorted by trial. Scripts were then written to take pertinent information for each analysis from this large initial structure. These scripts conducted the necessary computations to get data points of interest collapsed across sessions, morph lines, and eventually individuals. Finally, a last set of scripts was written to perform the statistical analyses referenced in my results.

Identification function. Investigating whether these data demonstrate the two main hallmarks of categorical perception previously mentioned, identification and discrimination, required two separate analyses. Identification is characterized by the aforementioned sigmoid or step function, the identification (ID) function (see Figure 1). Typically, the ID function results from plotting proportion of trials in which an image was identified or 'perceived' as one of the endpoint stimuli. The shape of these plots can be quantified by fitting different linear and nonlinear curve functions to the data and testing for the best fits. Given my y-axis represents proportions, curve fitting is not a statistically appropriate modeling tool, as it does not restrict points along the modeled curve to remain between 0 and 1 (as all proportion data points would). That said, a series of different curves were fitted to the data (linear, quadratic and exponential) to at least describe the overall shape of the response curve. A specific kind of regression analysis does exist to analyze whether a continuous variable (morph images along a morph line in this case) significantly predicts a categorical or binomial outcome variable (in this case the perception of endpoint stimulus one or two). This form of regression uses a sigmoid link function known as a logit function to transform the data and uncover a linear regression line between these two variables. Although I was not particularly interested in whether morph image significantly predicts perception per se, it does give a goodness of fit statistic for the logit function, which indirectly speaks to whether the ID function is sigmoidal. Therefore, a logistic regression analysis was performed on all the morph lines in this experiment. There is debate in the statistical community as to whether typical measures of goodness of fit for regression lines, such as R^2, are applicable to logistic regression lines (Peng, Lee, & Ingersoll, 2002). Peng and colleagues suggest that the best method is the Hosmer & Lemeshow test. It too has its limitations, the most considerable being that it is only a significance test and therefore only provides information about whether the fit is good or not. It does not provide insight as to how good (or poor) the fit is. Regardless, it is this statistic that I will report in this experiment. A significant result would indicate that the fit is not sound.

In MATLAB, a script was written to sort through all trials and parse them into sessions. For each morph image within a session, the number of 'left choices'- the number of times a subject 'perceived' the endpoint stimulus associated with the left checkerboard- was found. These data were then organized into morph lines, resulting in a structure with each element containing pertinent descriptors of the morph line along with the number of left choices made for each morph image across sessions. The mean response values were then plotted in a different script and fit to the aforementioned curves using the cftool in MATLAB. This script also performed a logistic regression on each of the morph line data and plotted the resulting fit.

Discrimination. The second hallmark of categorical perception, namely discrimination, is normally tested in a separate paradigm (Etcoff & Magee, 1992). Most studies, after collecting identification data and determining the category boundary across individuals, test a paradigm in which a separate sample of participants perform the discrimination task outlined earlier. Although I was not able to run a separate experiment to test for discrimination, I was able to do a proxy measure of the same phenomenon. In my data analysis, I compared mean response values for one morph image versus those for another. Two comparisons were made for each morph line- one that crossed the category boundary and another that fell to one side of the category boundary. The pairs of morph images that were being compared in either case were spaced equally along the morph line (see Figure 5). Placement of the category boundary can be difficult in data like these, but here I defined its position as the point along the morph line at which the mean response values crossed the 0.5 proportion mark (i.e., the transition point at which the subject stopped predominantly perceiving one endpoint image and began to predominantly perceive the other. This method of placement closely resembles that seen in the human literature (see Beale & Keil, 1995; Cheal & Rutherford, 2010), where the raw data, the ID function, is used to define the category boundary and not the curve fittings they used to assess the shape of the ID function. If stimuli were perceived categorically, one would expect the mean response values to be significantly different for the two images that straddle the category boundary but not so for the two images on the same side of category boundary. This comparison was evaluated using the following test statistic that transforms proportions to z scores:

$$Z = (\text{p hat}_1) - (\text{p hat}_2) \quad / \quad \surd(\text{p hat}(1\text{-p hat})\ (1/n_1 + 1/n_2))$$

Where,
p hat = $(Y_1 + Y_2) / (n_1 + n_2)$
p hat_1 = the first sample's proportion value
p hat_2 = the second sample's proportion value
n_1 = sample one's size
n_2 = sample two's size

Familiarity effect. Data will be analyzed to determine whether familiarity pulled the category boundary away from the familiar individual; a prediction based on the human categorization literature, which suggests that categorical perception is more clearly demonstrated when endpoint images are familiar, canonical images. In order to conduct analyses of the effect of familiarity, response curves across morph lines had to be lined up. A script was written to conduct both horizontal and vertical graphical flips of the ID functions such that the x and y axes of the familiar/unfamiliar ID functions lined up with one another.

Mean locations of the category boundaries in familiar/familiar versus familiar/unfamiliar response curves were compared using a Student's *t* test. Additionally, further analyses were conducted to measure qualitative differences between familiar/familiar and familiar/unfamiliar ID functions. One such qualitative difference was that familiar/unfamiliar ID functions appeared to have a more step-like, distinct ID function; a greater proportion of the mean response values were extreme. To statistically analyze this difference, mean response values were compared to their associated probability of occurrence using the binomial distribution for that particular sample size (n=18). The number of morph images with mean response values whose probability of occurrence by chance exceeded 0.001 (representing $p< 0.001$) was tallied and compared across familiarity conditions. It was thought that this analysis could be considered a measure of consistency in their perception of a certain morph image as belong to one identity or the other. Mean number of mean response values exceeding the aforementioned *p* value were compared across familiarity conditions using the aforementioned test statistic for comparing proportions. Upon inspection of the data, a second qualitative difference was noted. It appeared as though the category boundaries were more distinct. In other words, the number of mean response values around the category boundaries with chance level proportions were greater for familiar/familiar morph lines than those in familiar/unfamiliar morph lines. To explore this observation quantitatively, a Mann-Whittney *U* test was used to compare the mean number of chance level data points surrounding the category boundary across familiarity conditions.

Adaptation. Lastly, I was interested in looking at whether some sort of hysteresis inherent to the structure of the paradigm was causing a response bias. That is, whether endpoint trials directly preceding morph image trials were functioning as adapters, influencing the subject's perception of the morph image. There is a body of literature looking at adaptation effects in categorical perception (Daelli, 2011; Daelli, Rijsbergen, & Treves, 2010; Webster, Kaping, Mizokami, & Duhamel, 2004), giving reason for concern and validating exploration of this possibility in these data. A script was written in MATLAB creating a structure that pulled all morph trials that were preceded by an endpoint trial, along with the identity of the preceding endpoint trial and the choice behaviour of the morph trial. The number of 'matches' versus 'non-matches' with each of the two possible endpoint identities were then tallied in a separate structure for each morph image in each morph line. A 'match' was defined as a morph trial in which the choice behaviour indicated that the subject perceived the morph image to belong to the same identity as the preceding endpoint trial. It would follow that, if the identity of the preceding endpoint trial had no effect on the perceived identity of the current morph image, the proportion of 'matches' versus 'non-matches' for each identity would not significantly differ from chance. In this case, chance would be the mean response value for that morph image across trials. For example, if for a particular morph image endpoint stimulus one was perceived 14 out of 18 times, chance would dictate that the proportion of endpoint stimulus one matches (an endpoint one trial preceded the morph image) would also be 0.7778 (14/18). If the proportion of endpoint stimulus one matches for trials in which that morph image was presented significantly differed from 0.7778 (with a two-tailed alpha level of 0.05), presentation of endpoint stimulus one directly before the morph image was either pulling or repulsing the perception of endpoint one in the morph trial. This difference was analyzed using the following test statistic, which computes the probability that the proportion of matches for a given morph image belongs to a population distribution with a mean proportion identical to the mean response value:

$$Z = (\text{p hat}) - p_0 \ / \ \surd(p_0\,(1-p_0\,) \,/\, n)$$

Where,
p hat = the sample proportion value
p_0 = the expected proportion value dictated by chance
n = sample size

Results

Only individuals LU and AN reached the necessary thresholds to complete the training paradigm. LU was the first individual to successfully complete training and the first morph line tested with her was the familiar LE/familiar AN morph line (see Figure 8a). It was surprising to see the bias in her perception of LE versus AN. The ID function appeared more like that which had been predicted for morph lines in the familiar/unfamiliar condition. A different familiar/familiar morph line was run with her (familiar AN/familiar RI) to test whether this bias was the result of the particular stimuli themselves. This morph line produced a more centralized category boundary as expected (see Figure 9a), however, upon collecting AN's data, it appeared that the mere presence of a bias may not necessarily be the product of a familiarity effect. LU was then run on the planned familiar/unfamiliar morph lines. AN was only tested on the planned morph lines outlined earlier.

Identification function

For each morph line, mean response values were calculated and ID functions were plotted (see Figures 6a-12a). Contradictory to previous evidence from pilot studies and published literature, most ID functions in this study demonstrated a bias in the perception of one identity or endpoint image over another, regardless of familiarity. Category boundaries were not centralized, aside from the LU's familiar AN/familiar RI female ID function (see Figure 9a), and some of the analyses had to be adjusted to account for this shift. With regards to the analysis of the shape of the ID functions, this perceptual bias affected the goodness of the fit of the logit function in the logistic regression analysis. Adding extra pseudo data points on the stunted end of the ID functions did not improve the fit. Logistical regression analyses were nonetheless conducted on all ID functions (see Figure 13a for an exemplar fit) and Hosmer-Lemeshow statistics for these analyses are summarized in Table 2. Significant results were found for LU's familiar DO/unfamiliar male and familiar RI/unfamiliar female morph lines with the conservative alpha level of 0.01. Additional curve fits were conducted to test which best described the data and goodness of fit values are summarized in Table 3. Of these, quadratic and exponential functions fit the ID functions better than a linear function.

Category boundaries

An attempt was made to adhere to the original definition for the placement of the category boundary, where possible. There were two morph lines, however, for which this definition did not seem appropriate. These morph lines were looked at individually when determining the most appropriate placement of their respective category boundaries. LU's familiar AN/familiar RI ID function had two morph images with mean response values of 0.5 (see Figure 9b). In this case, it was decided that the category boundary should be placed in between the two. The other morph line in question was AN's familiar DO/unfamiliar male (see Figure 10b). Optimal placement of the category boundary in this case was slightly more ambiguous. It was decided that the greatest emphasis should be placed on the statistically significant mean response values and thus the mean response value associated with the 50% ‘familiar DO’ morph image was taken out of consideration. The category boundary was placed between the morph images that AN strongly associated with each of the two endpoint identities (see Figure 10b).

As previously mentioned, raw binomial data are used to define the category boundary in this study in accordance with previous practice in the literature. If one were to use the fitted logit functions to define the category boundaries for these ID functions, placement of the category boundaries would be shifted from the above defined boundaries into the flattened, within-category boundary area of the ID function and would not accurately represent the step in the function (number of steps shifted along morph line: μ=1.71, s= 0.7). Alternatively, fitting sigmoidal functions to the data and using these to define category boundaries would result in the same locations as those described above, with the exception of AN’s familiar DO/unfamiliar male morphline (see Figures 14-20), which is shifted in the direction of the unfamiliar female face and thereby demonstrating a bias in favour of familiar face DO.

Discrimination

Due to the perceptual biases seen in these ID functions, the selection of an appropriate reference morph image and interval length required careful consideration so as to accurately represent the ID functions obtained. It was decided that a two-step interval would be most appropriate, such that if the 50/50 morph was selected as the reference morph it would be compared to both the 30/70 morph and the 70/30 morph (with one of these morphs lying directly after the assigned category boundary and the other on the same side of the category boundary as the reference morph image). This placement accounted for the perceptual biases in most of these ID functions. The 'across-boundary' morph image was always the morph image closest to the category boundary on the side of the morph line associated with the endpoint identity least often seen in the morph images. This is graphically illustrated in Figure 5. For AN's familiar DO/unfamiliar male morph line, the interval was increased to a 3-step difference to account for the assumedly anomalous data point associated with morph image 50% 'familiar DO' (see Figure 10b).

For all morph lines except LU's familiar AN/familiar RI morph line, across-boundary comparisons of proportions were significantly different from each other and within-boundary comparison were not. These data are summarized in Table 4. In the case of LU's familiar AN/familiar RI female morph line, an inversed effect was observed. The across-boundary comparison was not significant, z=1.014, *p*= 0.1562, while the within-boundary comparison was, z=2.12, *p*= 0.0170.

Familiarity effect

Given the biases seen in the positioning of the category boundaries of most of the ID functions (see Figures 6b-12b), including those associated with familiar/familiar morphlines, the original hypothesized effect of familiarity had to be revisited. A significant shift in the position of the category boundary in ID functions associated with familiar/unfamiliar morph lines versus those associated with familiar/familiar morph lines was no longer demonstrable- the category boundaries of familiar/familiar ID functions were already shifted considerably in one direction such that little space was left to show a significantly greater shift in a given direction for familiar/unfamiliar ID functions, especially with such a small n (three and four, respectively). Although a shift in mean category boundary location was seen between ID functions associated with familiar/familiar morph lines (μ=3.1, sd=0.98, df=2) versus those associated with familiar/unfamiliar morphlines (μ=2, sd=1, df=3), this difference was not significant ($t(5)$= 1.33, sd= 0.83, $p > 0.05$). Additionally, the shift favoured the familiar endpoint stimulus in only three of the four familiar/unfamiliar ID functions. AN demonstrated a perceptual bias for the unfamiliar male over the familiar male, DO.

Visual comparison of familiar/familiar versus familiar/unfamiliar ID functions did, however, suggest qualitative differences between the two, namely the ID functions associated with familiar/unfamiliar morph lines appeared to be more demonstrative of distinct, categorical perception. This led to the additional analyses described earlier.

The mean number of mean response values that did not significantly exceed chance were compared across familiarity conditions (familiar/familiar: m=2.3, sd=1.528, df=2; familiar/unfamiliar: m=1.2, sd=0.5, df=3). Due to unequal variances and sample sizes of less than 10, a nonparametric Mann-Whitney U test was conduct on the means using an alpha level of 0.05 and the mean ranks did not differ significantly, U=11, crit value=0, $p > 0.05$. Cumulative number of mean response values not significantly exceeding chance for each of the familiarity conditions (familiar/familiar: 7/27; familiar/unfamiliar: 5/36) were also compared to see whether the proportions differed significantly using the aforementioned test statistic for comparing proportions. Again, the difference was not significant, $z = 0.52$, p=0.3015.

Lastly, the probability of obtaining each mean value by chance was computed and mean response values with probabilities that significantly exceeded chance at p values of 0.05, 0.01, and 0.001 were labeled according in Figures 6b-12b. A comparison of the proportion of mean response values exceeding $p<0.001$ across familiarity conditions was conducted. This analysis uncovered a statistically significant difference between the cumulative number of values exceeding p<0.001 in familiar/familiar ID functions (10/27) versus those in familiar/unfamiliar ID functions (25/36), z=-2.5617, p=0.0052. The same test was conducted to include data points exceeding a p value of 0.01, but this difference was no longer statistically significant with an alpha of 0.05, z= -1.2881, p=0.0985.

Adaptation

To evaluate whether the identities of endpoint stimuli were having an adaptation effect on the morph trials they preceded, the proportion of morph trials for each morph image in which the perceived ID of the morph image matched the ID of preceding endpoint stimuli were calculated. Z scores for all matches did not deviate significantly from that which would be expected by chance, aside from the match scores associated with the 10% 'familiar DO' morph image in LU's ID function (ID 'DO' match: $z= -2.0125$, $p= 0.0222$; ID 'unfamiliar male' match: $z= -2.4508$, $p= 0.0071$).

Discussion

Despite some variation in the data, these results demonstrate that rhesus macaques perceive conspecific faces in a categorical manner. Evidence of both hallmarks of categorical perception, namely identification and discrimination, was found when two rhesus macaques were asked to identify morphed images varying along a computer-generated morph continuum between two photographed conspecific faces. This study employed a two-forced-choice visual paradigm akin to those previously used to demonstrate categorical perception of conspecific faces in humans. This phenomenon becomes more distinct when one, but not both, of the conspecific faces presented in the morph line are personally familiar to the individual. These results appear to be the product of underlying neural mechanisms of visual perception that cannot be accounted for by response bias or inter-trial adaptation effects. Given the small sample size in this study and some of the unexpected idiosyncrasies of these data, further studies are required to strengthen the evidence found here and confirm whether the qualitative differences between the ID functions produced by subjects in this experiment and those presented in human literature can be generalized as interspecies differences. As these data are the first of their kind to be collected from non-human primates, no within-species comparisons are possible. Regardless, this study has laid down the behavioural foundation for future exploration of the neural underpinnings of the phenomenon known as categorical perception.

Although these data strongly suggest rhesus macaques perceive faces categorically, a few key characteristics of these data qualitatively separate them from those reported in human literature. The most notable departure was the perceptual bias evident in all but one (LU's familiar AN/familiar RI) of the ID functions collected from subjects- there is no indication from previous research that category boundaries of familiar/familiar morph lines should not be centralized on the morph continuum. The extreme values I report here could represent a fundamental, qualitative difference in categorical perception in human versus non-human primates. Given a previous study that demonstrated humans and rhesus macaques perform similarly on an object categorization task (Fabre-Thorpe et al., 1998), these differences are more likely attributable to subtle methodological differences that separate this study from those previously conducted with humans or individual differences amplified by a small sample size.

Although all attempts were made to model this study after the human literature, the face stimuli used here did vary from those in the literature with regards to familiarity. Familiar faces used in most human studies are canonical images of famous people that are assumedly acquired semantically through exposure to them in the media. Campanella, Hanoteau, Seron, Joassin & Bruyer (2003) tested categorical perception of personally familiar faces in humans using both an identification and discrimination task. The resulting ID functions mirrored the rest of the literature, reporting category boundaries falling between 42-58% along the morph line. That said, the 'familiar' images used in their study depicted professors in the same university program as the students tested and the morphlines were much shorter (5 morph images in length). The level of personal familiarity of conspecific faces used here is much greater and could be likened to that of immediate family members or long-term roommates. It is possible that judgments of perception may vary for faces we are exposed to and interact with on a daily basis. This difference in perceptual processing along with the use of a more sensitive 9-point morph scale could be responsible for the perceptual bias seen in this study. To test whether this differentiation between the personal relevance of the familiar face influences categorization, it would be interesting to try to implement more semantically acquired familiar faces with rhesus macaques. This could be achieved by displaying pictures of the faces in their living quarters and exposing them to videos of the same individual extensively over a long period of time before testing took place.

Beyond these speculations, it is important to note that this perceptual bias is not due to a simple response bias in choice behaviour. If these biases were attributable to a side bias alone, endpoint mean response values would not reach the accuracy levels they did and endpoint stimuli associated with the same checkerboard stimulus side would mirror the direction of bias in the data. LU would be expected to show the same biases for LE, AN, DO and the unfamiliar female face and AN would show the same biases for LU, DO and the unfamiliar female face, which simply was not the case. Additionally, adaption effects of preceding endpoint trial identity were explored and found to be largely insignificant (with the exception of one of 63 morph images presented showing significant results that could represent an adaptation effect for this particular morph). Regardless, more data would need to be collected before claims could be made with respect to fundamental differences in categorical perception between species.

Selection of an appropriate method to analyze whether the ID function was indeed sigmoidal proved difficult given the nature of these data. It is arguable based on visual inspection of the id functions that all, including the two morph lines that did not fit the logistic regression line, are indeed sigmoidal. The use of logistic regression in this case is problematic for a few different reasons. Firstly, my major interest was not to measure predictability, which is what logistic regression analyses are designed to do. The perceptual biases shift the ID functions drastically enough that the stunted end did not really allow for the logit function to truly capture the data. Elaborate curve fitting analyses would have been more sensitive to truly capturing the data and modeling the ID function. In retrospect, it would have been more appropriate to write sigmoidal functions for each in MATLAB, optimize them, and then test them for goodness of fit, such as those displayed in Figures 14-20. Overall, these data seem more like step rather than sigmoid functions. This would imply that the categorical perception reported was more distinct and therefore requires different statistical analyses than those conducted. Accordingly, fitting a step function to the ID functions and testing goodness of fit would be a prudent analysis to conduct. The curve fitting that was conducted on the ID functions was only intended as a tangential measure of best fit and was not expected to optimally model the data. Specifically, it was important to demonstrate that a linear fit does not model the data optimally. This was shown to be true for all ID functions arising from the morph lines tested.

Now that it has been demonstrated that stimuli of varying levels of complexity, from the chromaticity of light to conspecific faces, are perceived categorically, perhaps the most prudent next step in the exploration of categorical perception is to search for some sort of stimuli that are not perceived categorically. If they do exist, it could test the merit or validity of this paradigm; stimuli that are not perceived categorically should result in linear ID functions. If not, it would speak volumes about perceptual processing, supporting the concept that neurophysiology is fractal in nature; mechanisms underlying lower level processes are repeated in higher level processes. This concept is already demonstrated by the fact that low level stimuli, like light chromaticity, are categorically perceived in much the same manner as highly complex visual stimuli, like faces; some inherent mechanism occurs in perception that is best plotted with a sigmoid function and can be seen at various levels of perception, from detection of differences in light to categorizing complex objects such as faces. Attracter dynamics may play a key role in conceptualizing this common mechanism.

The discrimination measure used here, although not the same as used in other studies of categorical perception, possibly provides stronger evidence that these subjects perceived the conspecific face stimuli categorically, as it adopts a more binomial perspective when characterizing the step in perception from one face to the perception of another. In some ways, it measures whether the assigned category boundary truly represents a significant shift in perception. Results from this task were indicative of categorical perception for all but one ID function, LU's familiar AN/familiar RI ID function.

The ID function associated with this morph line was dissimilar to the others in many ways. Despite its category boundary being more centralized as seen in the human literature and although its fit with the logit function was considered 'good' by the Hosmer-Lemeshow test, it failed to display a discrimination effect, which may be considered a more hardy test of categorization in this case. This anomalous finding is difficult to account for. Considering the level of exposure to the individual LU demonstrated a perceptual bias towards in her LE/AN morph line, namely LE (with whom she has equal or less visual exposure to than to AN), and the fact that LU did not show a bias in RI's direction (the individual with whom she shares her living quarters) when paired with AN, this anomalous finding cannot be attributed to an effect of varying levels of familiarity. Again, due to a small sample size, I cannot rule out that this ID function's peculiar shape is simply due to noise in the signal.

Perhaps some of the most interesting results reported here regard the effect of familiarity on the ID functions. Although the initial hypothesis became difficult to test due to the perceptual biases found across familiarity conditions, statistically significant differences in the structure of the ID functions were found. To my knowledge, no previous study of human or non-human primates has investigated mixed-familiarity morph lines. Investigations of the effect of familiarity on categorical perception have always compared familiar face morph lines to unfamiliar face morph lines. Studies have suggested that familiarity results in a more distinct categorical perception (see Beale & Keil, 1995; but see Campanella et al., 2003). As such, it was hypothesized that providing one familiar stimulus would create a perceptual bias in favour of the familiar face. As the ID functions in this experiment from both familiarity conditions demonstrated perceptual biases and the sample size was extremely small, the necessary statistical power was lacking to uncover any significant effects in this perceptual bias between conditions. Regardless, the direction of the bias was consistent with my hypothesis in only three of the four familiar/unfamiliar morph lines. AN's familiar DO/unfamiliar male ID function demonstrated a perceptual bias in favour of the unfamiliar male. It is disconcerting that this shift was seen in the one morph line that was presented to both individuals. If this finding is not anomalous, it would imply that one cannot assume all individuals exposed to that same morph line would provide similar ID functions and calls into question whether the perceptual biases seen in these ID functions are the result of individual differences. Alternatively, if category boundary placement was determined by the fitted sigmoid functions seen in Figure 18, the bias would be in the predicted direction, favouring DO. Again, I am limited in my ability to draw conclusions about this finding due to the small sample size.

The significant difference across familiarity conditions in the consistency of identification was unexpected. If anything, the literature would suggest that pairing a familiar face with an unfamiliar face might introduce more noise into the data than that present in familiar/familiar pairings. Regardless, a clear, statistically significant difference was found describing the consistency of their perceptual appraisals of the morph image's identity. Familiar/unfamiliar ID functions resulted in more significantly-higher-than-chance mean response values at an alpha level of 0.001 than familiar/familiar ID functions and approached significance with values at an alpha level of 0.01. Although this is not consistent with what one might expect given the literature, it could be rationalized, albeit it feebly, from a biological evolutionary standpoint. One could argue that rapid discrimination between a familiar face and a complete stranger would facilitate an abrupt fight-or-flight response- we fear what we do not know. Additionally, it was argued based on familiarity effects reported elsewhere in the literature that categorical perception is not an innate property of the visual system but rather is reliant on heavy exposure to the stimuli. Findings from this study along with Campanella and colleagues' (2003) provide contradictory evidence to this hypothesis. Consistent perceptual appraisals were not limited to the side of the ID function associated with the familiar individual; unfamiliar individuals were consistently perceived in morph images to the left of the category boundary, as well.

It is important to preface further discussion of the implications of this study by acknowledging the difficulties associated with collecting behavioural data from an animal model, especially when a considerable amount of training is required before the animal can perform the experimental task. For this reason, sample sizes are generally fairly small, especially in non-human primate literature. Unfortunately, this study is no exception. Extensive efforts were made to collect data from four subjects, but limitations in time and trainability only allowed for collection from two individuals. Equally unfortunate is the fact that statistical power is linked to sample size- extremely small sample sizes as seen in this study limit its capacity to uncover effects, if present. For this reason, the major limitation of this study is its restricted capacity to allow for broad claims to be made based on the results found. Additionally, constraints in the pool of familiar individuals that could be pulled from necessitated the use of unique pairings across individuals. Ideally, all subjects should be shown the same morph lines of faces thought to be equally familiar to all subjects tested. For this reason, the only strong claim I can make is that categorical perception of conspecific faces can be demonstrated in rhesus macaques using similar methods as those seen in humans. Broad claims about the quality of that categorical perception cannot be made.

These limitations considered, the validity of this study remains sound and its impact factor is still considerable. Now that behavioural data has been collected confirming categorical perception of conspecific faces in rhesus macaques, neural studies can follow, elucidating key properties of the neural mechanisms responsible for this perceptual phenomenon. As is the case with most of our knowledge about neurophysiological systems in humans, further exploration is dependent on the use of an appropriate animal model. Studies have collected evidence regarding different brain areas that may implicate their importance in the categorical perception of faces, as well. The inferior temporal cortex (ITC) has been shown to play a role in object recognition and categorization (Freedman, Riesenhuber, Poggio, & Miller, 2003; Kiani, Esteky, Mirpour, & Tanaka, 2007; Mruczek & Sheinburg, 2007; Sigala, 2004; Wilson & Debauche) in tandem with prefrontal cortex (PFC; Freedman et al., 2002, 2003). The superior temporal sulcus (STS) may be responsible for certain aspects of categorization (Linden, Turennout, & Indefrey, 2009) and encoding of facial expression (Furl, Rijbergen, Treves, Friston and Dolan, 2007). Face discrimination has been associated with activity in the fusiform face area (FFA; Dotan, Gelbard-Sagiv, & Malach, 2009) and, given its recognition component, researchers have explored the role different medial temporal structures play in rapid categorization (Fize, Boulanouar, Chatel, Ranjeva, Fabre-Thorpe & Thorpe, 2000). By demonstrating that rhesus macaques perceive faces in a categorical manner, this study validates future electrophysiological studies investigating how these different regions together are responsible for face categorization in the brain.

These neural findings would have clinical relevance, as well. One study has already demonstrated that otherwise high functioning adolescents with autism show significant impairments in face categorization (Teunisse & Gelder, 2001). Neural findings following this study could be used firstly as a biomarker of the condition and secondly as a clue to what may be going on in the autistic brain- a question currently on the minds of many scientists and granting agencies around the world. Although this study on its own provides but a small piece to the behavioural puzzle known as categorical perception, it marks a significant step in the direction of understanding mechanisms common to many different levels of perceptual processing and opens the door to future electrophysiological studies with wide-reaching scientific and clinical implications.

References

Beale, J.M., & Keil, F.C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217-239.

Bornstein, M.H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287-300). New York: Cambridge University Press.

Bornstein, M.H., & Korda, N.O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research, 43*(3), 207-222.

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*, 81-118.

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. *Visual Cognition, 10*(2), 129-156.

Cheal, J.L, & Rutherford, M.D. (2010). Mapping emotion category boundaries using a visual expectation paradigm. *Perception, 39*, 1514-1525.

Daelli, V. (2011). High level adaptation aftereffects for novel objects: The role of pre-existing representations. *Neuropsychologia, 49,* 1923-1927.

Daelli, V., Rijsbergen, N. J. van, & Treves, A. (2010). How recent experience affects the perception of ambiguous objects. *Brain Research, 1322,* 81-91.

Dahl, C.D., Logothetis, N.K., & Hoffman, K.L. (2007). Individuation and holistic processing of faces in rhesus monkeys. *Proceedings of the Royal Society B, 274,* 2069-2076.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorization of natural scenes does not rely on colour cues: A study of monkeys and humans. *Vision Research, 40,* 2187-2200.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology, 23(*1), 1-22.

Etcoff, N.L., & Magee, J.J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227-240.

Fabre-Thorpe, M., Richard, G., & Thorpe, S.J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport, 9*(2), 303-308.

Fize, D., Bouanouar, K., Chatel, Y., Ranjeva, J-P., Fabre-Thorpe, M., & Thorpe, S. (2000). Brain areas involved in rapid categorization of natural images: An event-related fMRI study. *NeuroImage, 11,* 634-643.

Fize, D., Cauchoix, M., & Fabre-Thorpe, M. (2011). Humans and monkeys share visual representations. *PNAS, 108(*18)*,* 7635-7640.

Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2003). Visual categorization and the primate prefrontal cortex: Neurophysiology and behaviour. *J Neurophysiol, 88*, 929-940.

Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience, 23*(12), 5235-5246.

Frisch, K. von. (1964). Bees: Their vision, chemical senses, and language. Ithaca, NY: Cornell University Press.

Furl, N., Rijsbergen, N.J. van, Treves, A., Friston, K.J., & Dolan, R.J. (2007). Experience-dependent coding of facial expression in superior temporal sulcus. *PNAS, 104*(33), 13485-13489.

Gelder, B de, Teunisse, J.P., & Benson, P.J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition & Emotion, 11,* 1-23.

Gilaie-Dotan, S., Gelbard-Sagiv, H., & Malach, R. (2010). Perceptual shape sensitivity to upright and inverted faces is reflected in neuronal adaptation. *NeuroImage, 50,* 383-395.

Girard, P., Jouffrais, C., & Kirchner, C.H. (2008). Ultra-rapid categorization in non-human primates. *Animal Cognition, 11,* 485-493.

Girard, P., & Koenig-Robert, R. (2011). Ultra-rapid categorization of fourier-spectrum equalized natural images: macaques and humans perform similarly. *PLoS ONE, 6*(2), 1-13.

Kiani, R., Esteky, H., Mirpour, K, Tanaka, K. (2007). Object category structure in response patterns of neural population in monkey inferior temporal cortex. *J Neurophysiol, 97,* 4296-4309.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54,* 358-368.

Linden, M. van der, Turennout, M. van, Indefrey, P. (2009). Formation of category representations in superior temporal sulcus. *J Cog Neurosci*, *22*(6), 1270-1282.

Mace, M.J.M., Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Spotting animals in natural scenes: efficiency of humans and monkeys at very low contrasts. *Animal Cognition, 13,* 405-418.

Martin-Malivel, J., & Okada, K. (2007). Human and chimpanzee face recognition in chimpanzees (pan troglodytes): Role of exposure and impact on categorical perception. *Behavioural Neuroscience, 121*(6), 1145-1155.

Mruczek, R.E.B., Sheinberg, D.L. (2007). Activity of inferior temporal cortical neurons predicts recognition choice behaviour and recognition time during visual search. *J Neurosci, 27*(11), 2825-2836.

Parr, L.A. (2011). The inversion effect reveals species differences in face processing. *Acta Psychologica, 138,* 204-210.

Peng, C-Y. J., Lee, K.L., Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3-14.

Perrett, D.I., & Emery, N.J. (1994). Understanding the intentions of others from visual signals: neurophysiological evidence. *Cahiers de Psychologie Cognitive, 13,* 683-694.

Ramon, M., Caharel, S., & Rossion, B. (2011). The speed of recognition of personally familiar faces. *Perception, 40,* 437-449.

Rotshtein, P., Henson, R.N.A., Treves, A., Driver, J., & Dolan, R.J. (2004). Morphing marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*(1), 107-113.

Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behavioural Brain Research, 149*, 1-7.

Teunisse, J.P., & Gelder, B de. (2001). Impaired categorical perception of facial expressions in high functioning adolescents with autism. *Child Neuropsychology, 7,* 1-14.

Tiddeman, B., Stirrat, M., & Perrett, D. (2005). Towards realism in facial transformation: results of a wavelet MRF method. *Computer Graphics Forum, Eurographics conference issue*, Vol 24, No 1-5.

Webster, M.A., Kaping, D., Mizokami, Y., Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*, 557-561.

Wilson, M., & DeBauche, B.A. (1981). Inferiotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia, 19,* 29-41.

Young, A.W., Rowland, D., Calder, A.J., Etcoff, N.L., Seth, A., & Perrett, D.I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271-313.

Preliminary Training Period	**Test Day 1**	**Test Day 2**	**Test Day 3**
Training Session(s)	**Booster Training**	**Booster Training**	**Booster Training**
	(until exceeds 85% accuracy)	(until exceeds 85% accuracy)	(until exceeds 85% accuracy)
Presented:	**Test Session 1**	**Test Session2**	**Test Session 3**
endpoint stimuli (120 trials/set)	Presented:	Presented:	Presented:
Criterion:	54 morph	54 morph	54 morph
85% accuracy	306 end point	306 end point	306 end point

Table 1. *Testing paradigm.* Structure of both the training and testing phases of the paradigm for each morph line presented.

Morph line	Chi-square	df	Significance
LU			
Fam DO/unfam	35.358	7	0***
Fam RI/unfam	19.32	7	0.007***
Fam LE/fam AN	9.315	7	0.231
Fam AN/fam RI	4.755	7	0.69
AN			
Fam DO/unfam	13.302	7	0.502
Fam LE/unfam	11.248	7	0.129
Fam LU/fam RI	14.967	7	0.036*

Table 2. *Hosmer-Lemeshow statistics from linear regression analyses for each morph line.* A significant result indicates an ill fit of the logit function to the ID function listed. $^*p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

				Goodness of fit	
Subject	**Morph line**	**Model**	**R^2**	**df**	**R^2- adjusted**
LU	Fam DO/ unfam male	Linear	0.6134	9	0.5705
		Quadratic	0.9351	9	0.9188
		Exponential*	0.9605	8	0.9562
LU	Fam RI/ unfam female	Linear	0.6173	9	0.5748
		Quadratic	0.8901	9	0.8626
		Exponential*	0.9264	8	0.9182
LU	Fam LE/ fam AN	Linear	0.772	9	0.7467
		Quadratic	0.9011	9	0.8764
		Exponential*	0.9126	8	0.9029
LU	Fam AN/ fam RI	Linear	0.9459	9	0.9399
		Quadratic*	0.9598	9	0.9497
		Exponential	0.9219	8	0.9133
AN	Fam DO/ unfam male	Linear	0.8487	9	0.8319
		Quadratic*	0.8839	9	0.8549
		Exponential	0.8458	8	0.8287
AN	Fam LE/ unfam female	Linear	0.5646	9	0.5162
		Quadratic	0.8663	9	0.8329
		Exponential*	0.936	8	0.9288
AN	Fam LU/ fam RI	Linear	0.7818	9	0.7576
		Quadratic*	0.9274	9	0.9093
		Exponential	0.8986	8	0.8873

Table 3. ***Goodness of fit statistics for curve fits.*** *Model with best fit (highest R^2)

Morph lines	Across-boundary comparison		Within-boundary comparison	
	z score	*p* value	z score	*p* value
Subject LU				
Fam/unfam male	3.2563	0.006***	1.0142	0.1562
Fam/unfam female	4.34	<0.00003***	0.603	0.2743
Fam/fam (LE/AN)	1.8091	0.0351*	-0.4211	0.3372
Fam/fam (AN/RI)	1.0142	0.1562	2.1213	0.0170*
Subject AN				
Fam/unfam male	-3.7187	<0.0001***	-0.603	0.2743
Fam/unfam female	3.1229	0.0009***	0	0.5
Fam/fam (LU/RI)	3.6723	<0.0002***	0.4211	0.3372

Table 4. ***Across- versus within- boundary comparisons of proportion values for morph images separated by two morph levels on their respective morph lines.***
*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

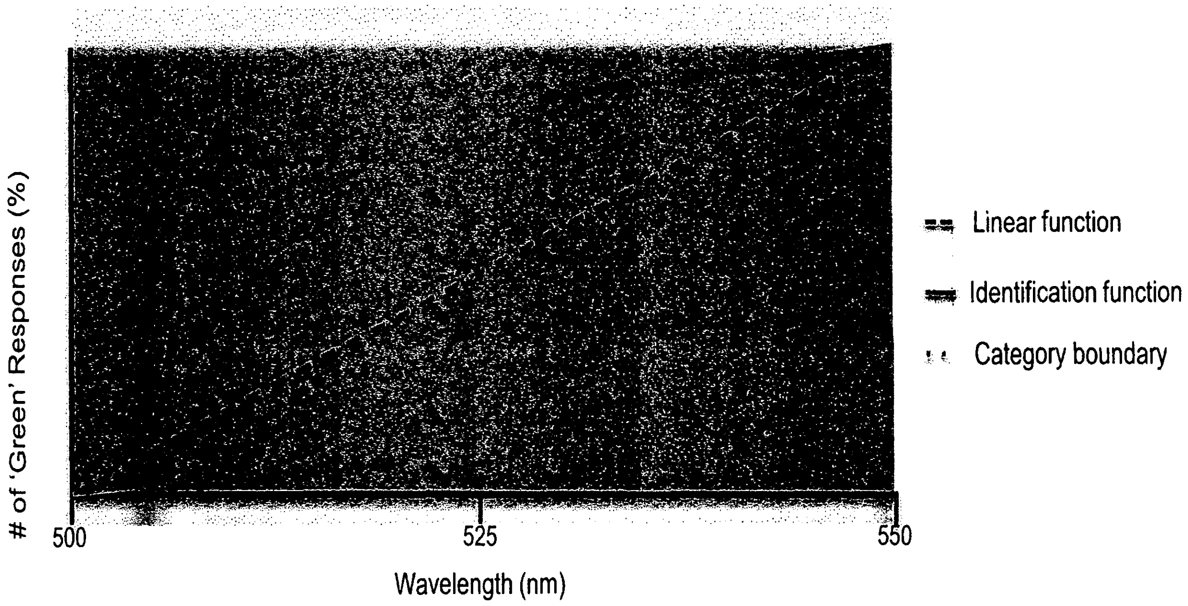

Figure 1. ***Exemplar identification (ID) function for hue.***

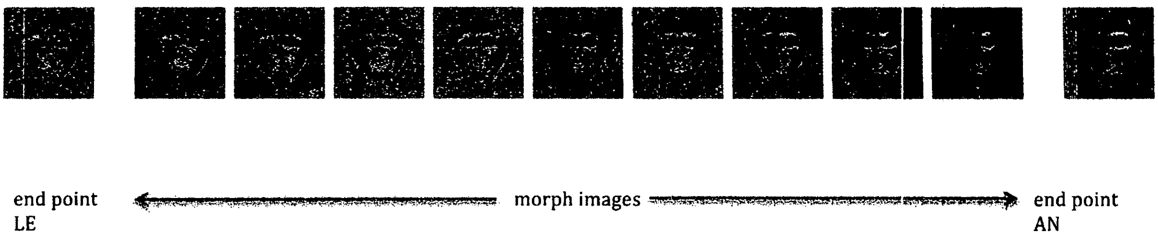

Figure 2. ***Exemplar morph line.*** Familiar LE/familiar AN morph line presented to subject LU with nine morph images along a continuum between the two endpoint images.

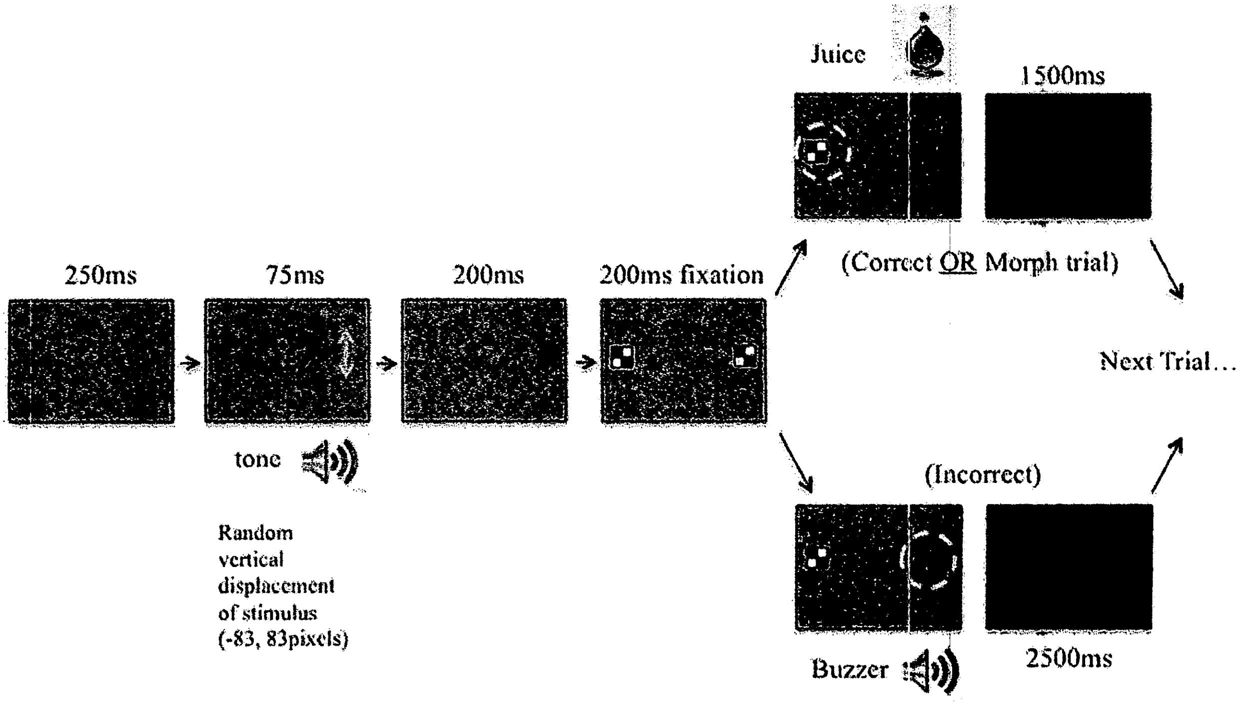

Figure 3. ***Trial Structure.*** Each square represents what is viewed on the computer screen by a subject for each phase of the trial. The yellow arrows are not presented on the screen, but rather are shown here to indicate varying vertical displacement of the face stimuli across trials. While dashed-line circles are also not presented on the screen, but are included here to indicated fixation of the subject.

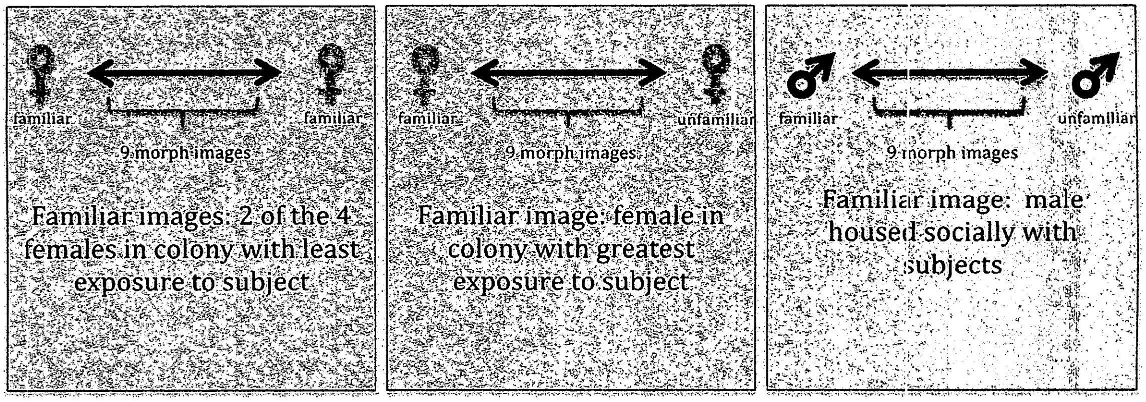

Figure 4. ***Morph lines presented to subjects.*** Each square represents one of the three 9-point morph lines planned to be shown to each subject.

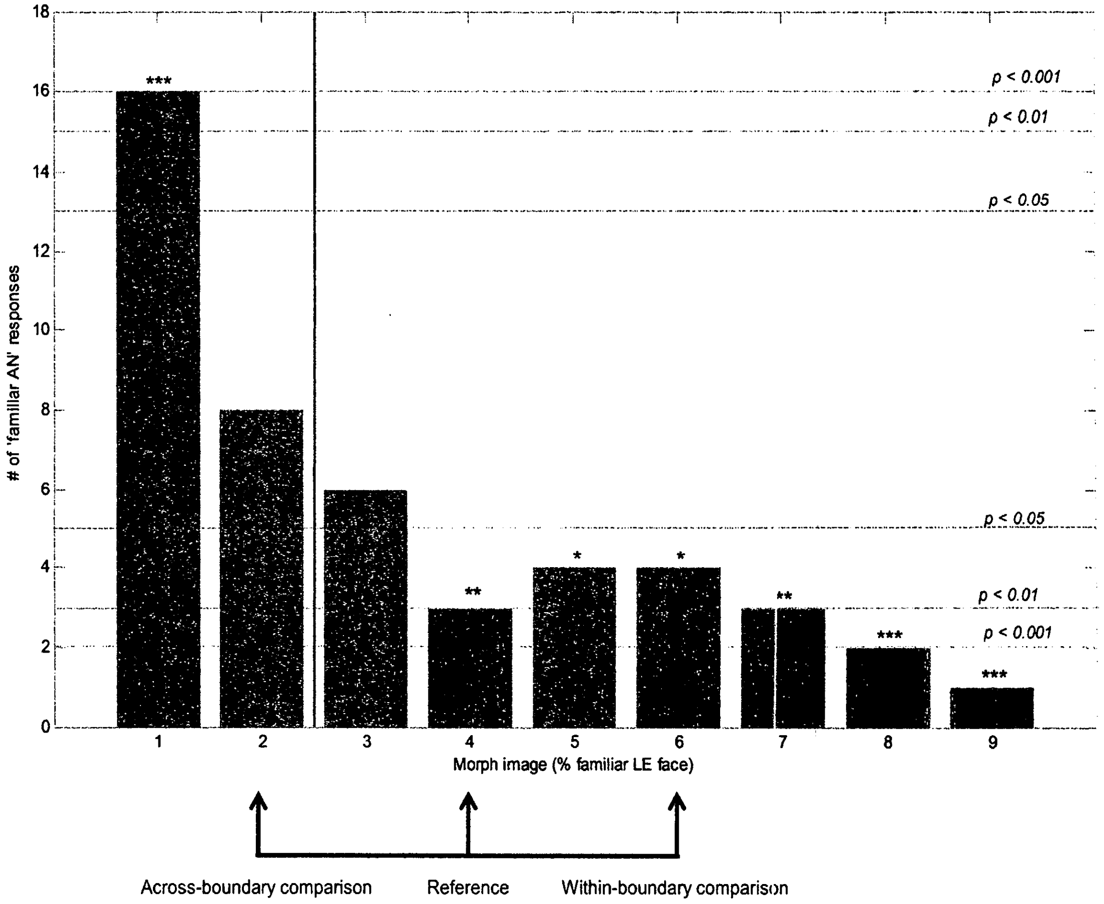

Figure 5. ***Method of selection of morph images for discrimination analysis.*** Arrows indicate which morph images' mean response values were compared in the discrimination analysis. The reference morph image's mean response value was compared to the across-boundary morph image's mean response value and the within-boundary morph image's mean response value in separate statistical analyses.

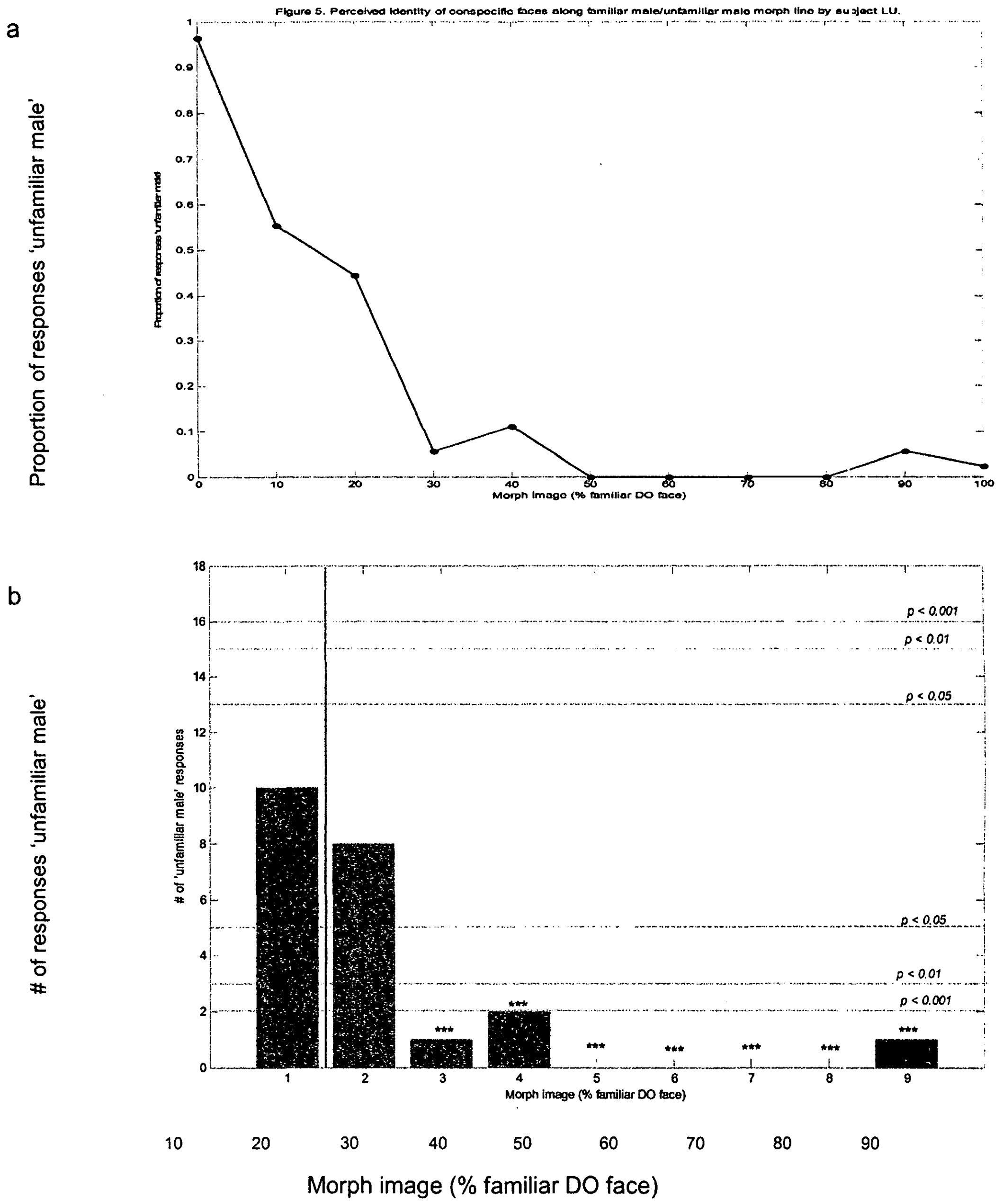

Figure 6. ***ID function produced by LU's responses to familiar DO/unfamiliar male morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. *$p<0.05$ **$p<0.01$, ***$p<0.001$

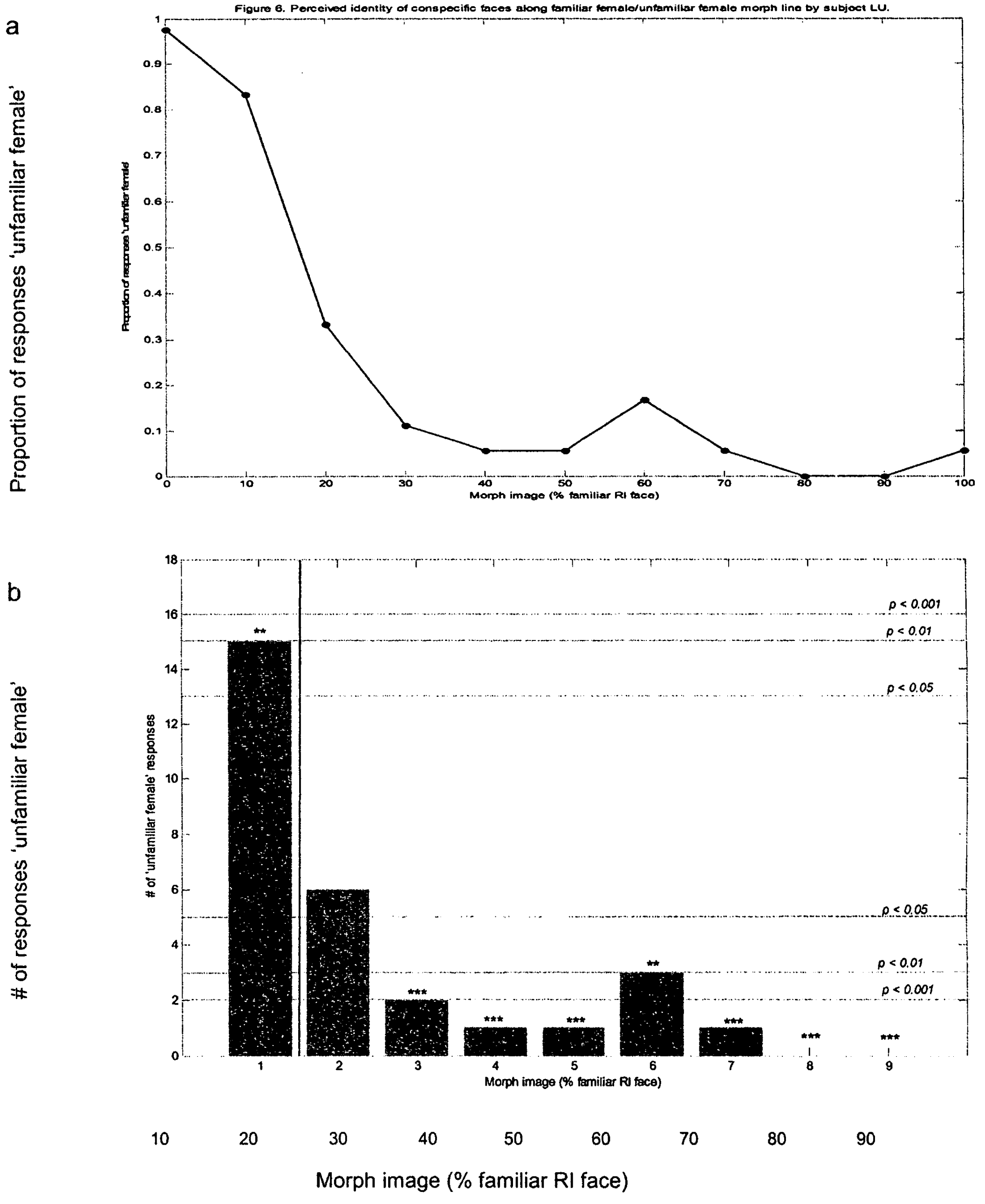

Figure 7. ***ID function produced by LU's responses to familiar RI/unfamiliar female morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. *$p<0.05$ **$p<0.01$, ***$p<0.001$

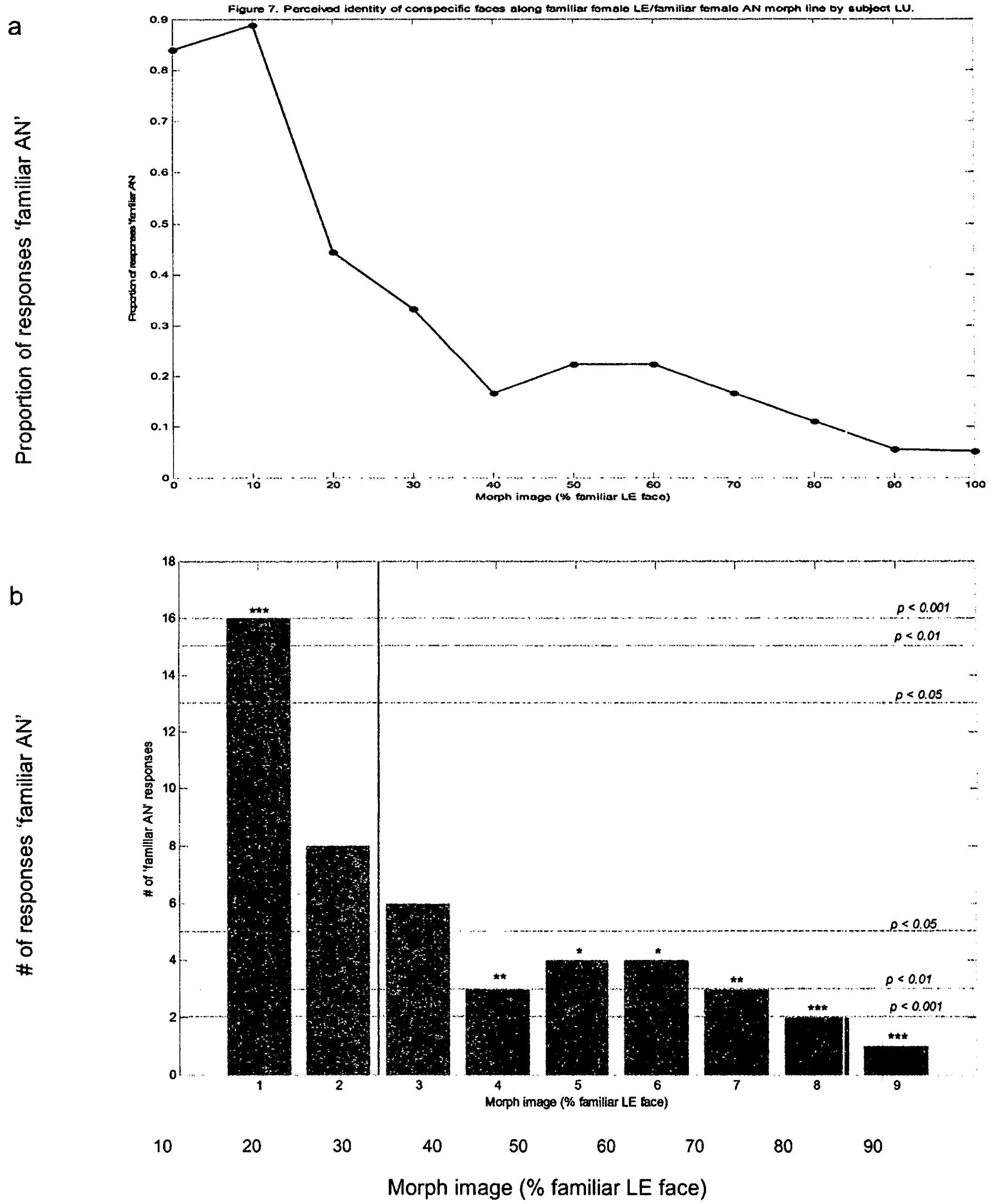

Figure 8. ***ID function produced by LU's responses to familiar LE/familiar AN morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. *p<0.05 **p<0.01, ***p<0.001

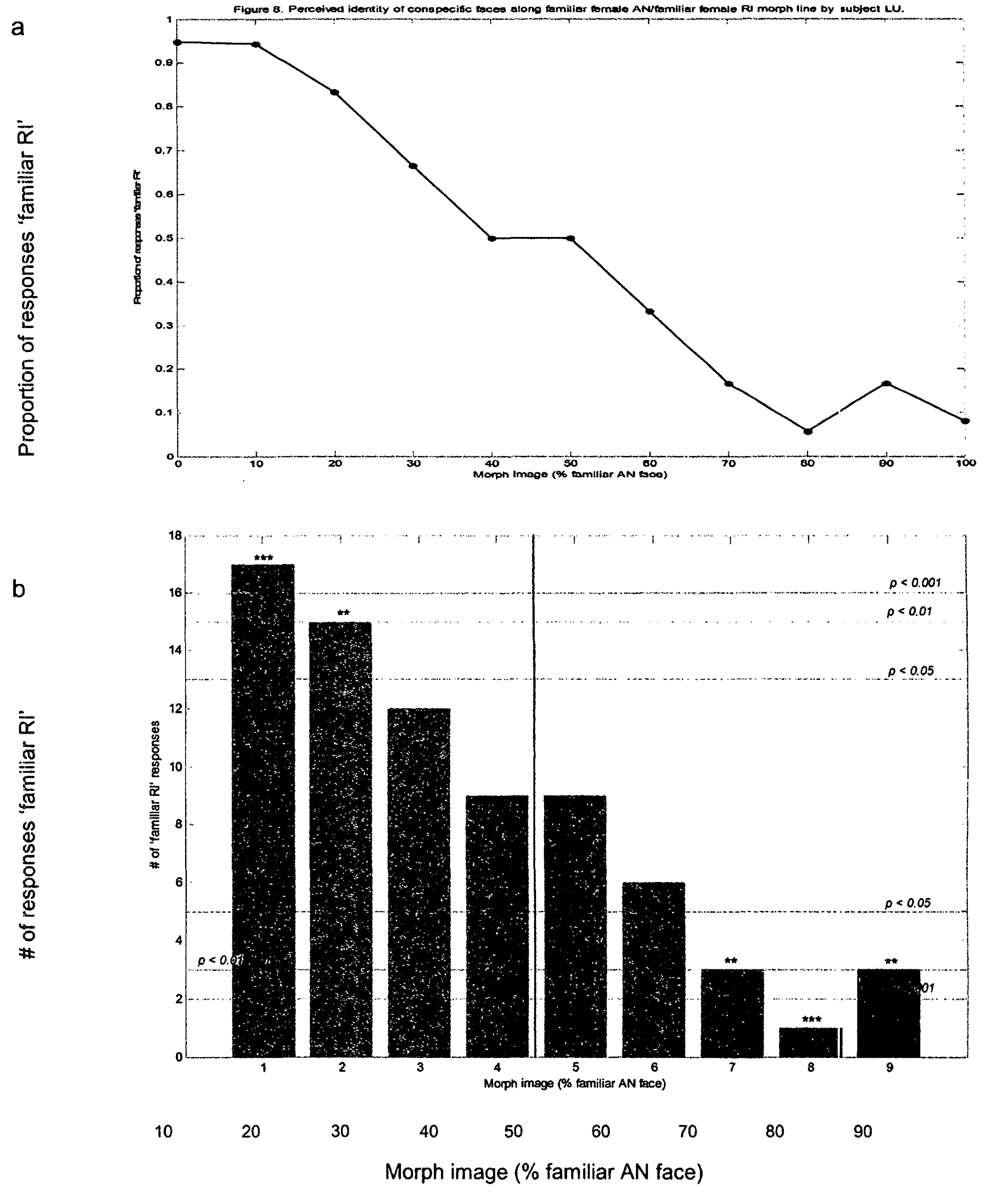

Figure 9. ***ID function produced by LU's responses to familiar AN/familiar RI morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. **p*<0.05 ***p*<0.01, ****p*<0.001

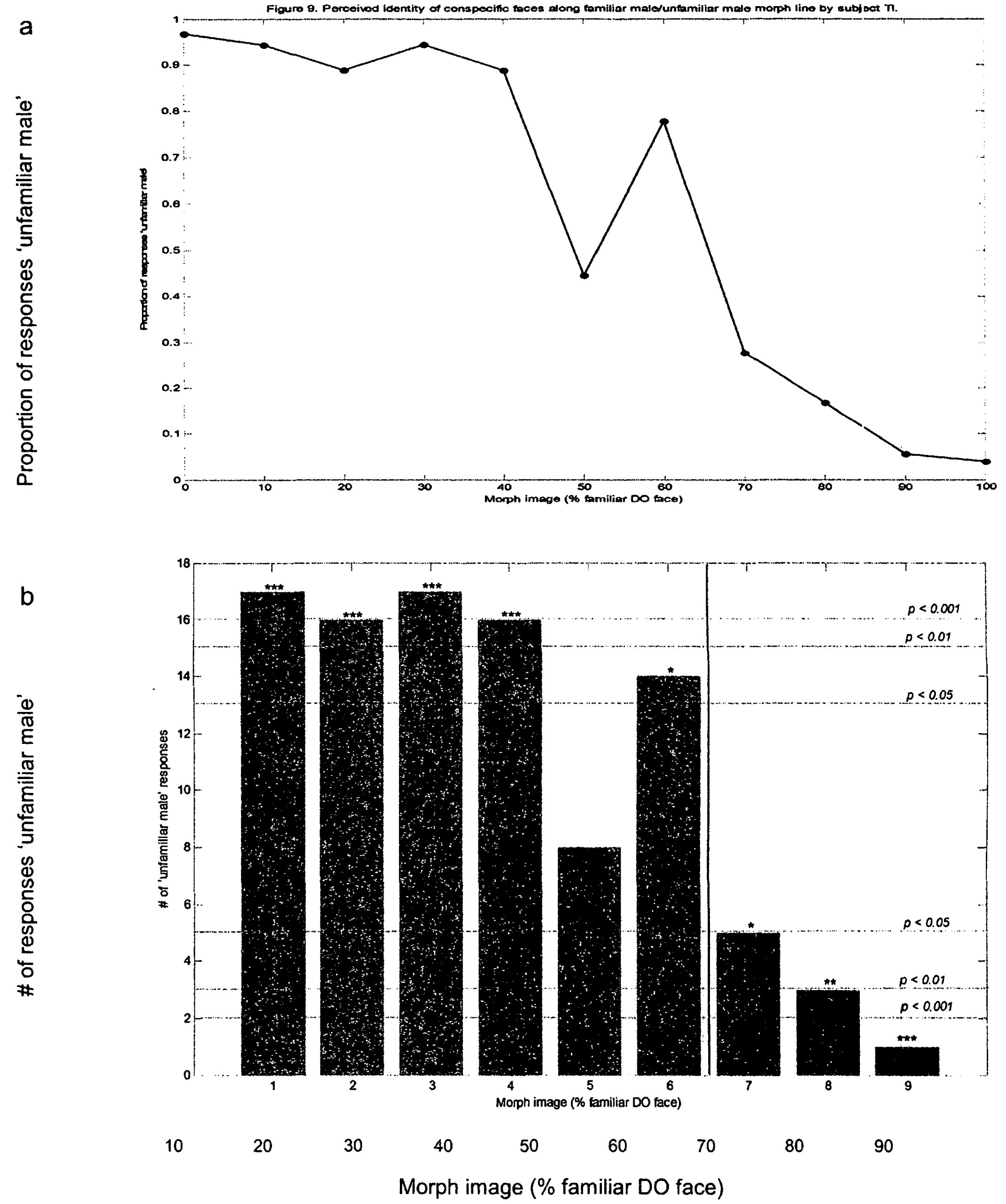

Figure 10. ***ID function produced by AN's responses to familiar DO/unfamiliar male morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. *$p<0.05$ **$p<0.01$, ***$p<0.001$

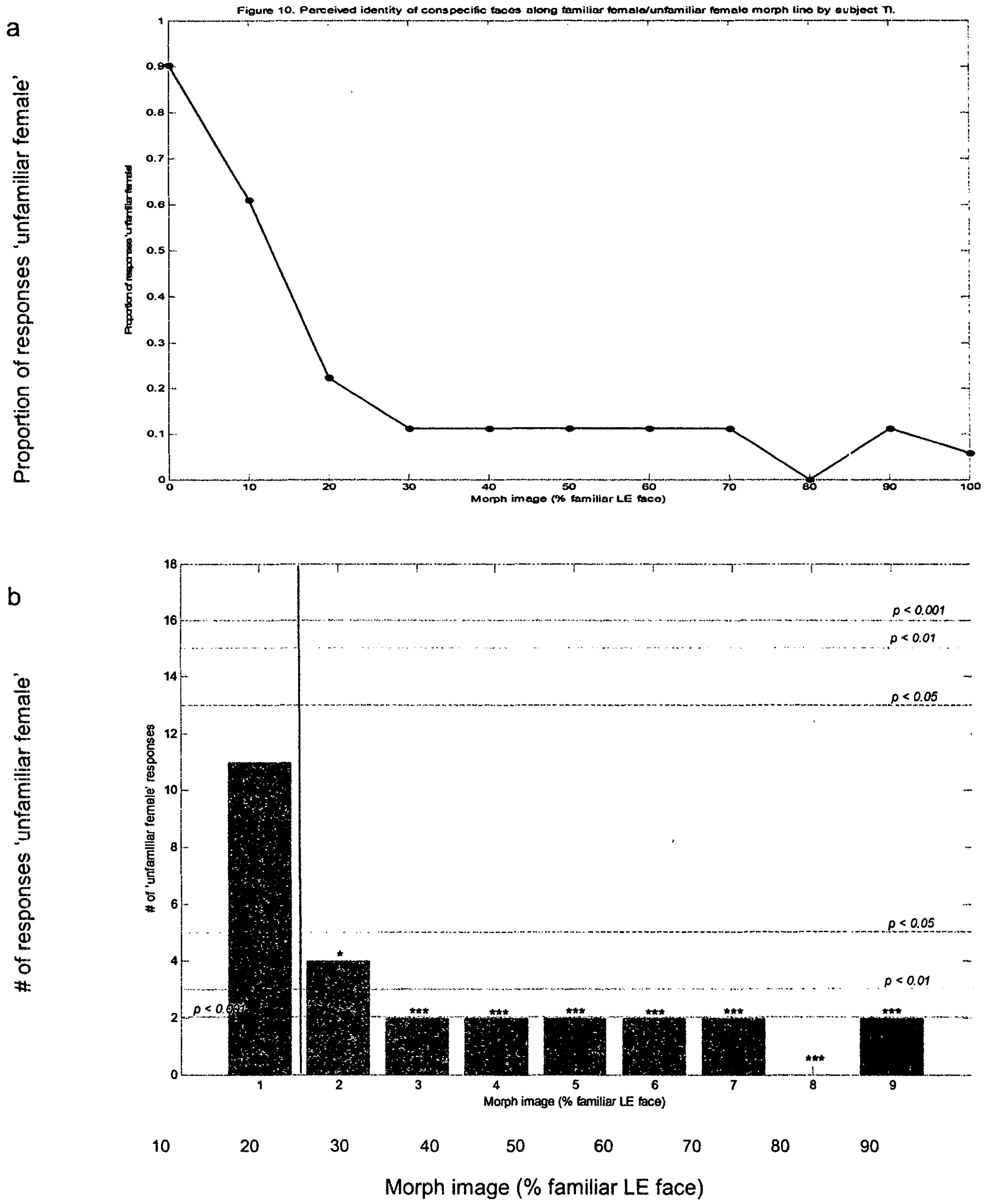

Figure 11. ***ID function produced by AN's responses to familiar LE/unfamiliar female morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. $^*p<0.05$ $^{**}p<0.01$, $^{***}p<0.001$

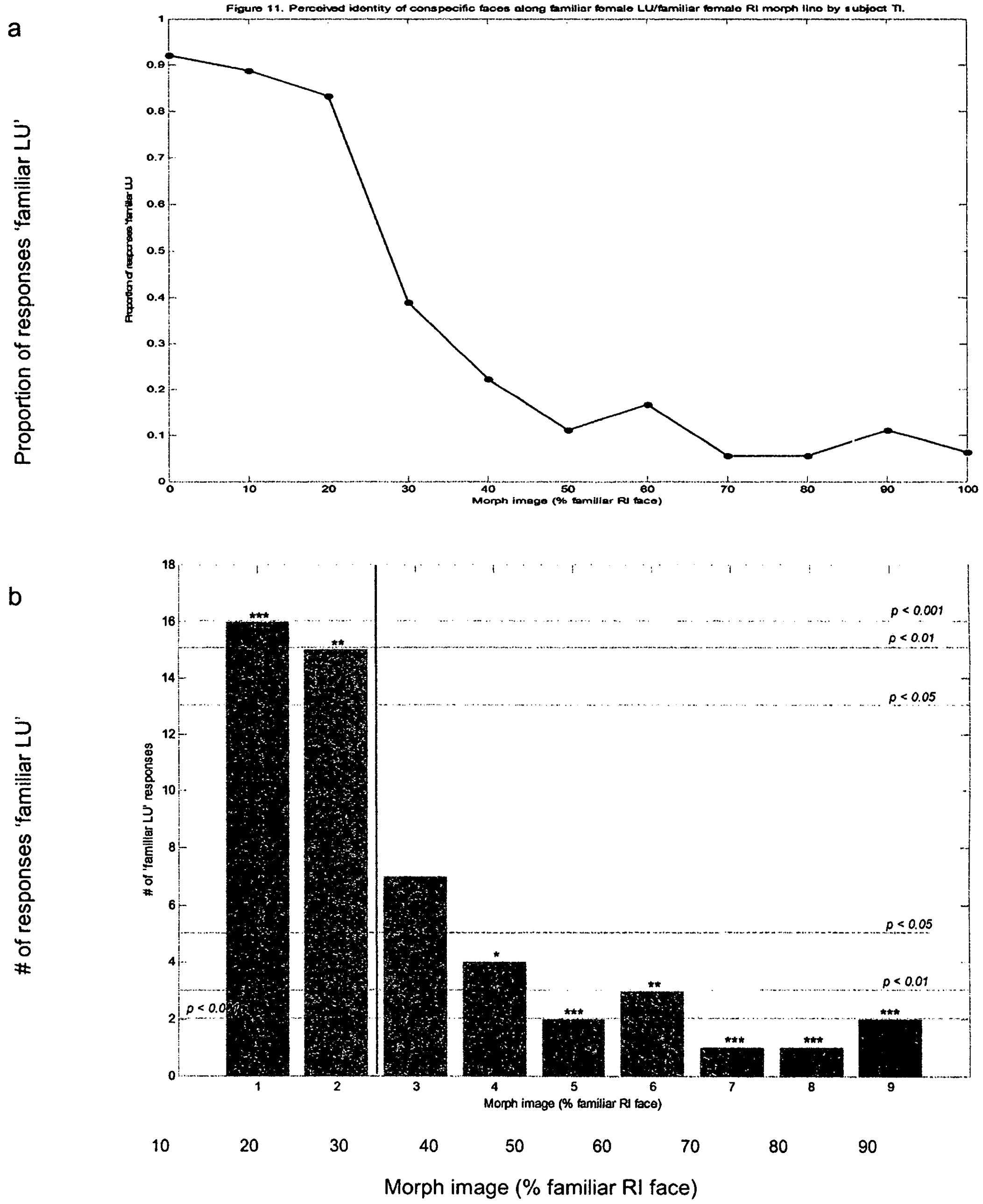

Figure 12. ***ID function produced by AN's responses to familiar LU/familiar RI morph line stimuli.*** **a** Line plot of mean response values used in logistic regression analysis **b** bar graph of proportional data with significance levels and demarkated category boundary used in descrimination and familiarity analyses. **p*<0.05 ***p*<0.01, ****p*<0.001

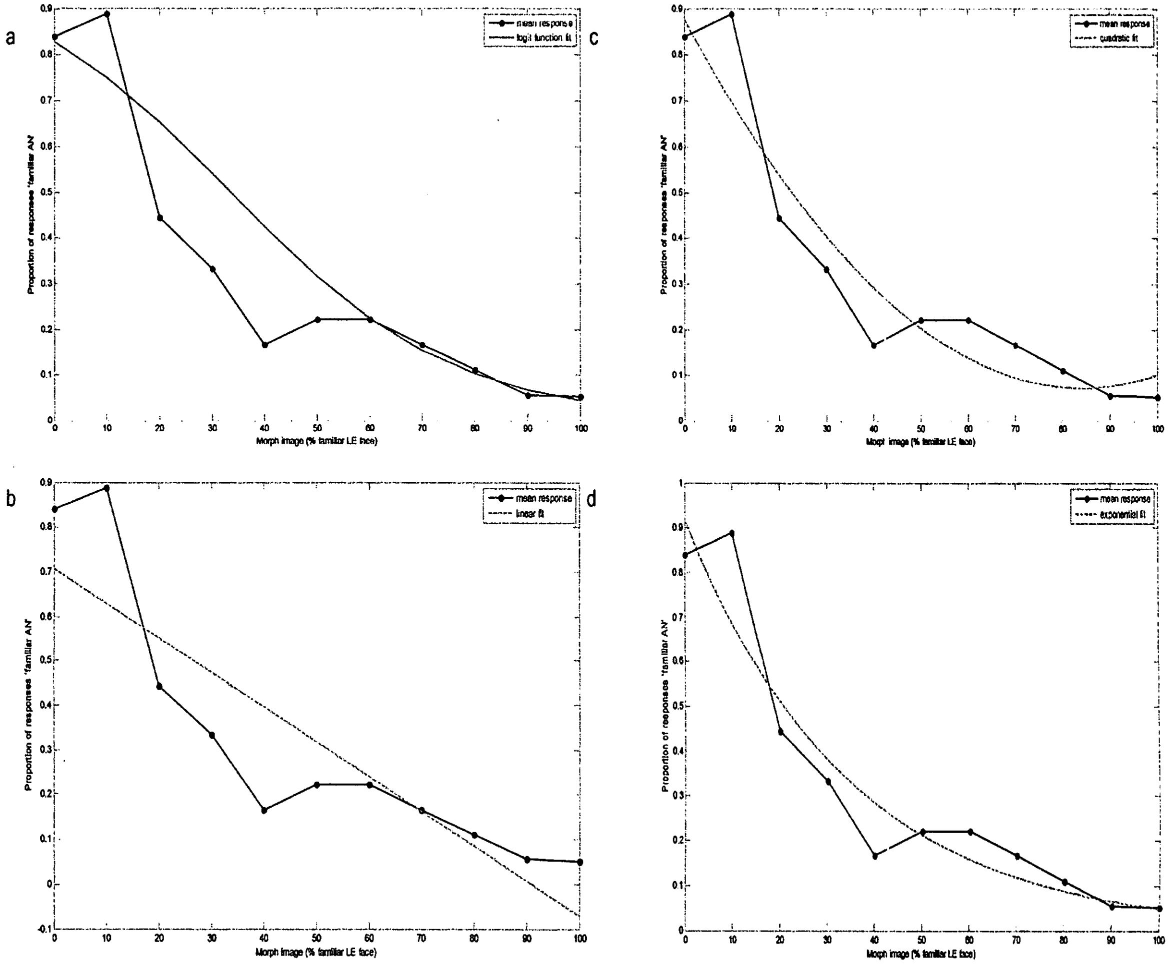

Figure 13. ***Example of the curve fits and logistic regression line computed for each ID function (LU's familiar LE/familiar AN morph line).*** **a** Logistic regression model **b** linear model **c** quadratic model **d** exponential model

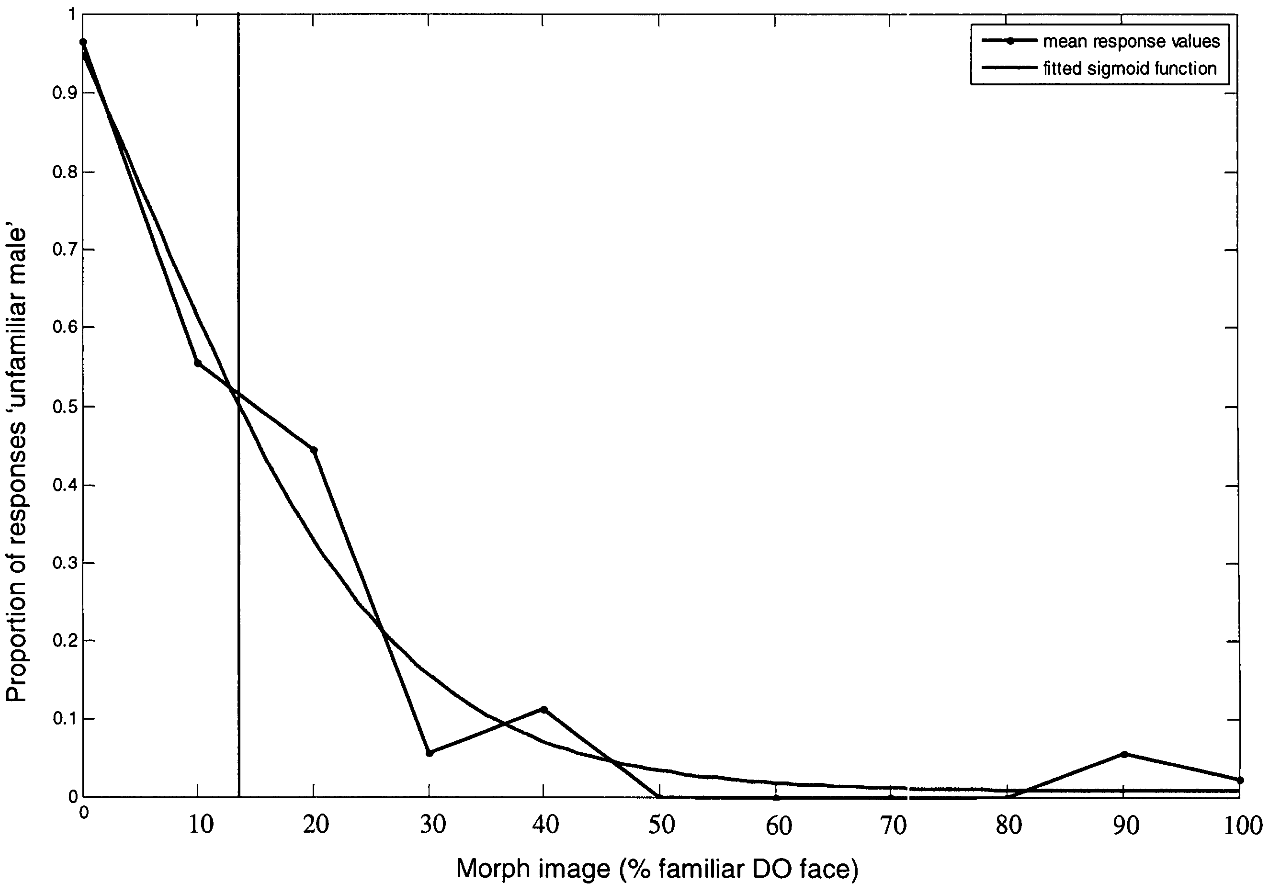

Figure 14. ***Alternative sigmoid function fit for LU's familiar DO/unfamiliar male morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.

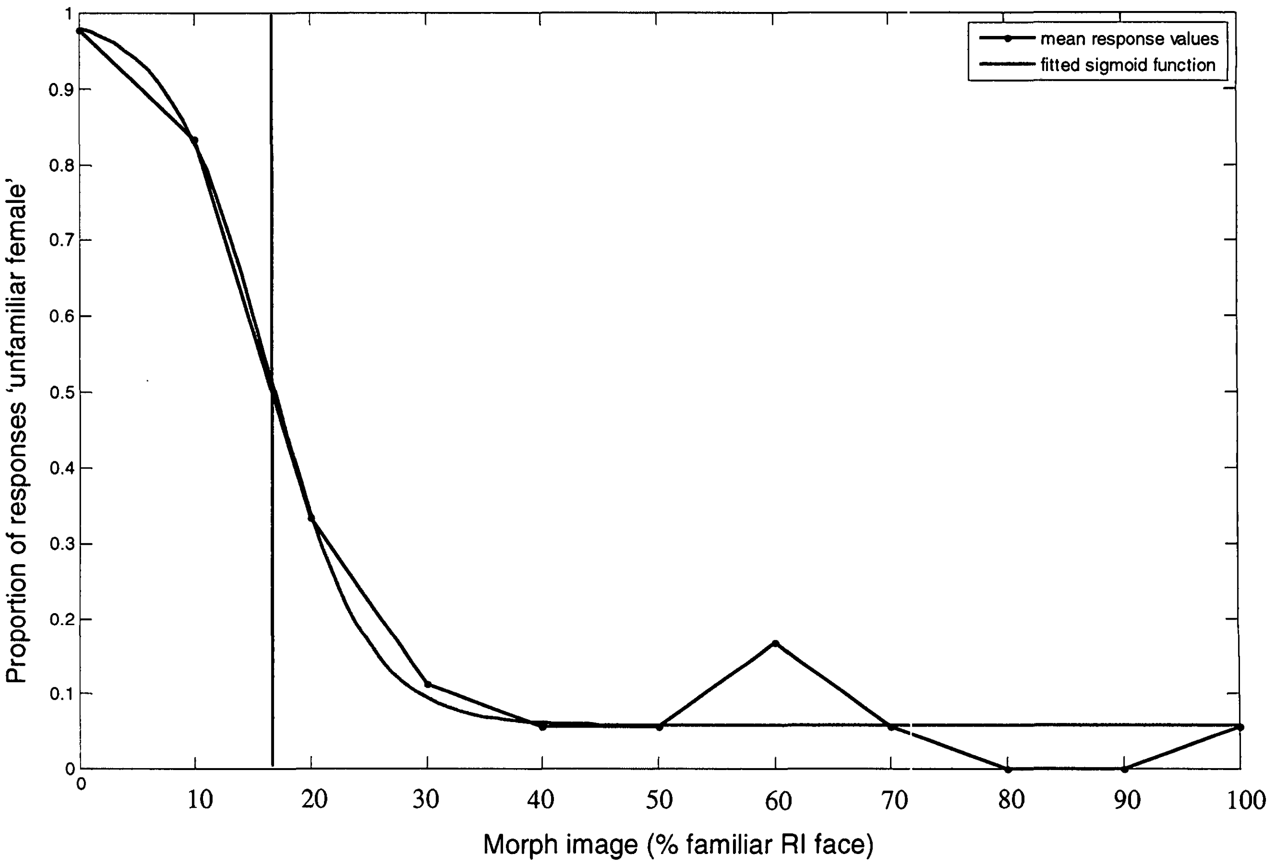

Figure 15. ***Alternative sigmoid function fit for LU's familiar RI/unfamiliar female morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.

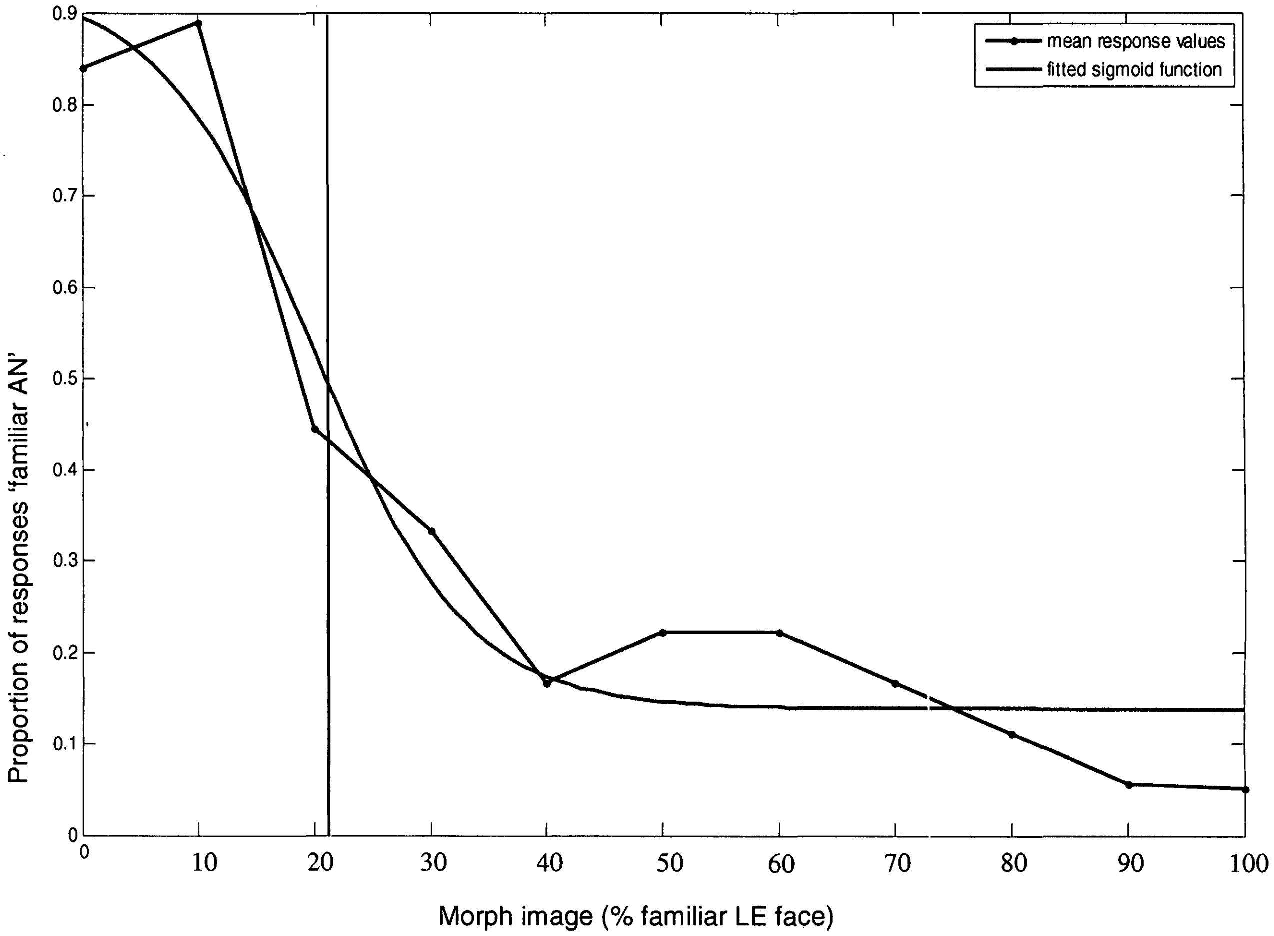

Figure 16. ***Alternative sigmoid function fit for LU's familiar LE/familiar AN morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.

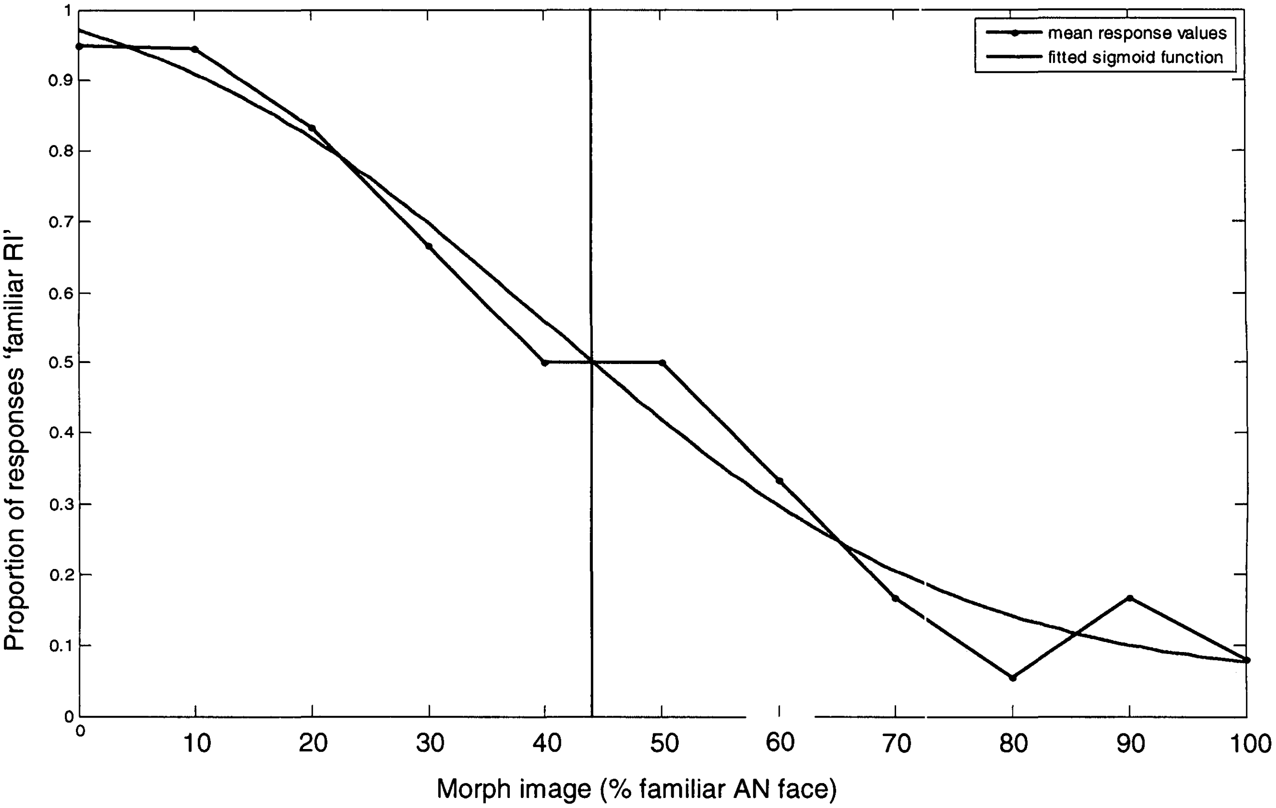

Figure 17. ***Alternative sigmoid function fit for LU's familiar AN/familiar RI morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.

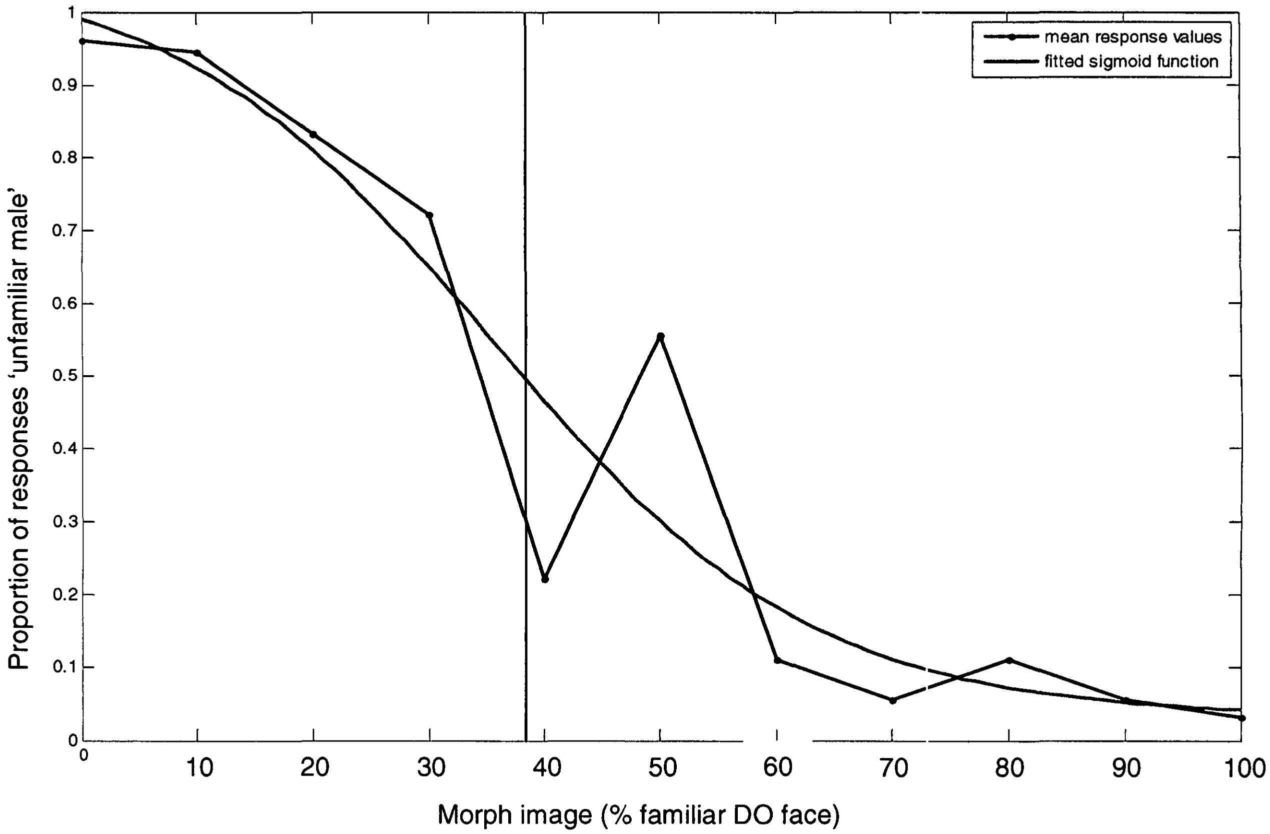

Figure 18. ***Alternative sigmoid function fit for AN's familiar DO/unfamiliar male morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.

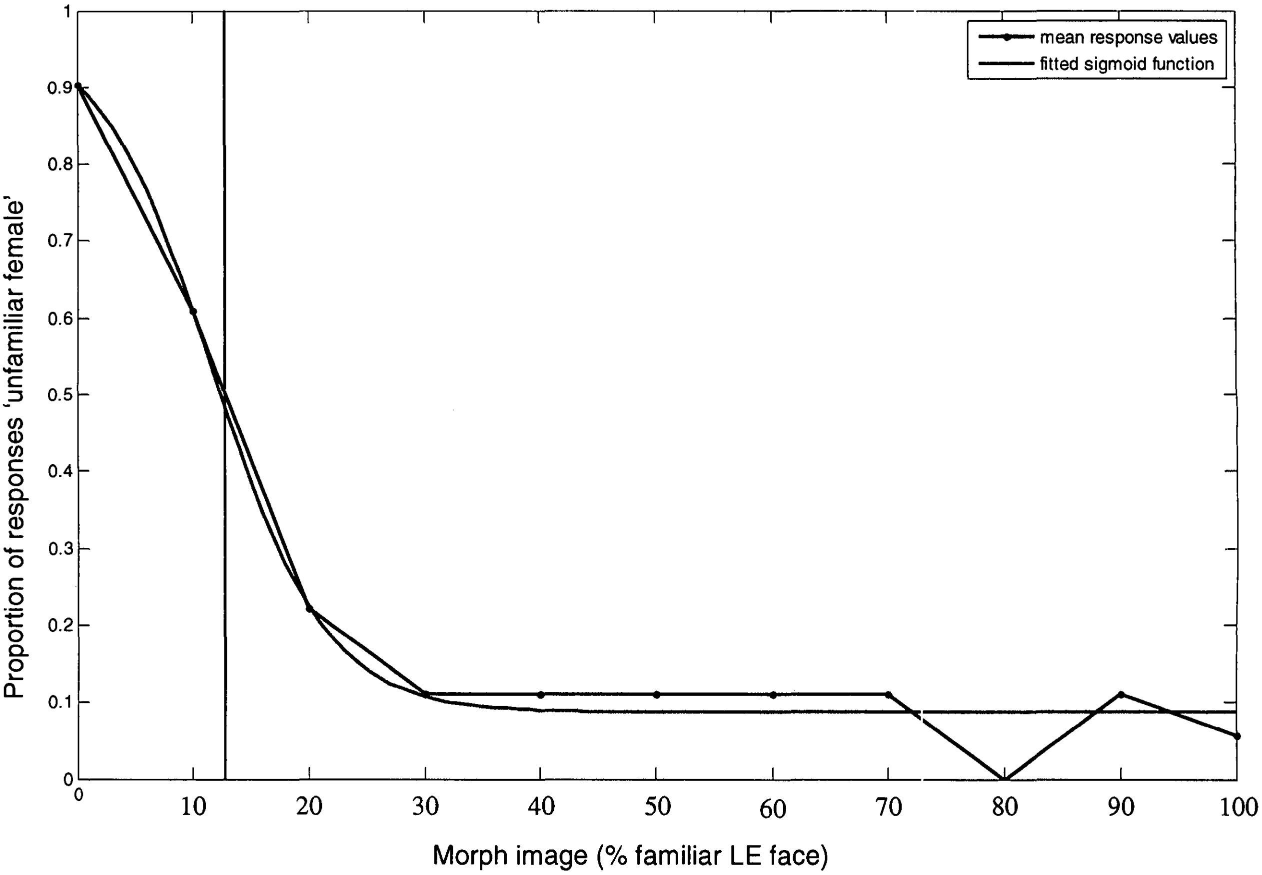

Figure 19. ***Alternative sigmoid function fit for AN's familiar LE/unfamiliar female morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.

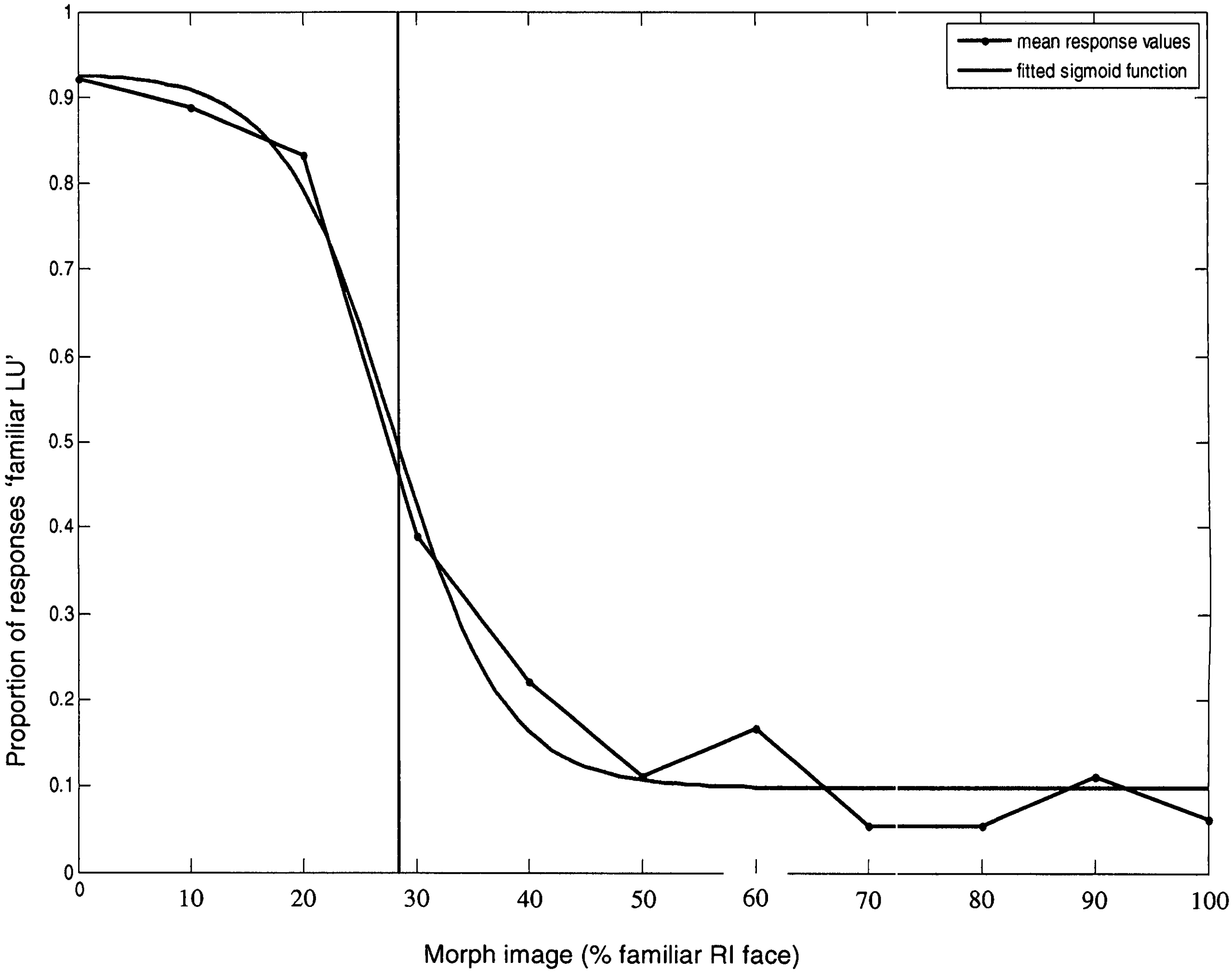

Figure 20. ***Alternative sigmoid function fit for AN's familiar LU/familiar RI morph line.*** The solid black line indicates placement of the category boundary, if determined by this sigmoid fit.