EMERGING HOLISTIC PROPERTIES AT FACE VALUE: ASSESSING CHARACTERISTICS OF FACE PERCEPTION

Mario Fifić

Submitted to the faculty of the University graduate School
In Partial fulfillment of the requirements
For the degree
Doctor of Philosophy
In the Department of Psychology and Cognitive Science,
Indiana University
January, 2006

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

James Townsend, Ph.D

Richard Shiffrin, Ph.D

Jerome Busemeyer, Ph.D

Daniel Maki, Ph.D

September 8, 2005

To Biljana and Lenka.

Acknowledgment

I want to thank to my committee, Dr. James T. Townsend, Dr. Richard Shiffrin, Dr. Jerome Busemeyer, and Dr. Daniel Maki, for their help in designing, analyzing and writing this dissertation. I also want to thank Dr. Michael Wenger for helpful suggestions; and thanks to Leslie Blaha, Dr. Ami Eidels and Chris Honey for fruitful discussions, as well as their help on thesis preparation and help with data collection.

The patient participation of my subjects is gratefully acknowledged.

Mario Fifić

Emerging holistic properties at face value:

Assessing characteristics of face perception

Holistic face recognition refers to the ability of human cognitive systems to deal in an integrative manner with separate face features. A holistic mental representation of a face is not a simple sum of face parts. It possesses unitary properties and corresponds to the whole face appearance better than to any of its constituent parts. A single face feature is better recognized in the learned face context (e.g. Bill's nose in Bill's face) than in isolation or in a new face context (e.g. Bill's nose in Joe's face; Tanaka & Sengco, 1997). The major goal of this study is to provide a rigorous test of the structure and organization of cognitive processes in the holistic perception of faces. Participants performed in two types of face categorization tasks that utilized either a self-terminating or an exhaustive rule for search (OR and AND conditions). Category membership was determined by the manipulation of two configural properties: eye-separation and lips-position. In the first part of each study, participants learned two groups of faces, and we monitored the changes in the face recognition system architecture and capacity. In the second part, the participants' task was to recognize the learned configurations of face features, presented in different face contexts: in the old learned faces, in a new face background and in isolation. Using the systems factorial theory tests,

combined with statistical analyses and model simulations, we were able to reveal the exact organization of the mental processes underlying face perception. The findings supported a view that holism is an emergent property which develops with learning. Overall, processing exhibited a parallel architecture with positive interdependency between features in both the OR and AND conditions. We also found that face units are better recognized in the learned face condition than in both the new face context and isolation conditions. We showed that faces are recognized not as a set of independent face features, but as whole units. We revealed that the cognitive mechanism of positive dependence between face features is responsible for forming holistic faces, and provided a simulation that produced behaviors similar to the experimental observations.

TABLE OF CONTENTS

LIST OF TABLES

The “OR” condition

The “AND” condition

LIST OF FIGURES

Faces are one of the most important aspects of our perception from many points of view. A wide variety of information is available when perceiving faces: they convey very important social messages by communication and expressing emotions, and thus provide additional information such as personality, social class level, age, health, gender, etc. It is not surprising then that face perception is an area of utmost importance in psychology and in other disciplines, including machine learning and clinical psychology. It is impressive with how little effort face recognition occurs. It is a very fast process, effortless, highly accurate and obligatory - you can not look at a familiar face and avoid recognizing it (see Young & De Haan, 1995).

Another striking aspect is the so-called non-homogeneity problem: unlike object between-class recognition, all faces share the same basic configuration. Relatively high discriminability between numerous faces is achieved only by small displacement of face-features so that extremely reliable discrimination occurs within the class of face-objects. It is therefore believed that face perception is a different form of perception, distinct from the perception of other visual objects (Bruce & Humphreys, 1994; Diamond & Carey, 1986; Farah, Levinson, & Klein, 1995; Farah, Rabinowitz, Quinn, & Liu, 2000; Farah, Wilson, Drain, & Tanaka, 1998; Moscovitch, Winocur, & Behrmann, 1997; Tarr, 2003; Tarr & Cheng, 2003; Yin, 1969). Support is provided on neuro-anatomical level by showing exclusive brain areas activated only during face perception. The fusiform face area (FFA) is preferably active to faces than for non-face objects (Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher, McDermott, & Chun, 1997); event-related potentials (ERPs) of the visual cortex showed selective responses for faces exhibited by peak negative activation at about 170ms after stimulus onset (Bentin, Allison, Puce, Perez, &

et al., 1996). Evidence against brain face-modularity is suggested by findings that FFA activity is correlated with expertise, even with non-face objects (*Greebles*; Gauthier, Skudlarski, Gore, & Anderson, 2000); in neural cell recording studies with macaque monkeys, non-face object selective neurons were found that are strikingly similar to the face selective ones (see also Gauthier & Curby, 2005; Haxby et al., 2001). As we could see in the previous examples the psychology of face perception generated large scale interest of scholars, from different areas (behavioral, cognitive, computational, neural etc).

The major goal of this study is to provide a rigorous test of the structure and organization of the cognitive processes in - what is defined as - holistic/configural perception of complex stimuli such as faces. Before we explicitly define holistic and configural we will use the terms interchangeably. A second important aspect of this study is that it utilizes experimental manipulations of faces so that they appear in more realistic perceptual context and therefore encourage more natural holistic/configural processing.

Realistic faces

In the first part of this study manipulations used to test cognitive organization are based on realistic stimuli. Here a more realistic face refers to the use of whole-faces rather than isolated face parts, as well as manipulations of faces that are close to their real life appearance – such as a blurred picture or partially occluded face. The standard approach to test processing of configural perceptions is based on comparisons of whole-faces to partial-faces (presentation of single face features while other are removed) (Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999; Farah, 1992; Homa, Haver, &

Schwartz, 1976; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Further, we will provide evidence regarding utilization of face parts together with whole-faces affects recognition of holistic aspects. We will refer to this property as blocking and mixing conditions.

"Face task"

In addition to previous manipulations which are employed with tests for architecture and capacity, another part of the study will be focused on a task that emphasizes recognition of a whole face. *Whole face* is defined as a complex face stimuli consisted of all face parts which are perceived together. Although the notion of whole face sounds like a pleonasm, in the experimental sense it has a different meaning. In many investigations concerned with explanation of configural and holistic face processing, *part of a face* is used (feature is presented in isolation such as a picture of lips or a nose only) and compared to processing of a whole face (all face features are presented).

We will use only whole faces that are partially occluded by a transparent mask over some important features on a face. While this is not a new manipulation, it has never been used together with a rigorous set of tests for processing architecture and capacity. The critical task instruction is to determined face category membership regardless of other possible aspects such as identity, emotion and gender. At the same time this manipulation will render some stimuli that are very similar to appear in some realistic scenarious, but the main advantage is to test the *whole-face* perception – where we believe the strongest holistic processing effect that could be elicited possibly lies.

Holistic and analytic processing of faces

First, we will approach the definitions of both holistic and analytic in more general terms, and then through the first part of this paper we will provide a stronger, more detailed definition, comparing it to previous approaches and to a large body of scientific evidence.

Holistic, well-configured, configural and Gestalt are all adjectives used to describe perception of a complex stimulus that is not perceived as a sum of individual features but rather as one perceptual element. Faces are visual stimuli that could be described as possessing this holistic property. But holistic perception is not necessarily limited only to faces. Gestalt psychologists believed that one of the fundamental principles of visual perception is determined by the “good form” of the perceived objects (Koffka, 1935).

Consider the following theoretical example: in a crowded place you are trying to locate a familiar friend’s face. Because of his/her familiarity, you need only limited information to recognize her/him. Let’s assume that you are limited to search for your friend’s eyes only, and that you are relatively successful in recognizing him. Also, assume that in another situation you have to search for two facial features at the same time, such as particular eyes and a nose. The question is: does adding more search features yield slower or faster processing?

It has been evidenced that performance declines when perceptual load increases, in both memory (Sternberg, 1966) and visual search (e.g. Atkinson, Holmgren, & Juola, 1969). These ubiquitous findings were usually backed up the assumption of capacity limitation. If our cognitive system is limited capacity, we could expect processing to slow down when we increase the total amount of information to be processed. But what happens in our example when you are looking for a familiar face in a crowd? Studies have been conducted where the number of face features were manipulated (Innes-Ker, 2003). In contrast to typical findings in memory and visual search, performance during face perception is facilitated when the information load increases. It was demonstrated that presenting a face with more face features (eyes, nose, lips) produced faster recognition of that whole face expressing a particular emotion (happiness, anger, etc.) (Innes-Ker, 2003). How does that happen? Why is face perception different than other object perception? Two general directions of have been taken over the course of the years. One was defined as analytic and the other one is holistic processing.

Analytic perception assumes that all face features are processed independently: a face is visually segregated into several elements such as eyes, nose, lips, ears, chin, etc. Each segment is perceived separately, suggesting that in our cognitive system there exist separate face feature detectors. It also suggests that a face is memorized as a set of different memory traces for each independent feature which are conjointly identified.

On the other hand, **holistic** processing assumes that a face is processed as a unitary stimulus instead of set of independent features. A face is perceived as a whole, not as a set of elements. Therefore, a face will make a single memory trace representation.

Both approaches can serve the findings reasonably well (will be discussed in the next section in more details). While is it relatively clear what it is meant by analytic processing, it is not intuitive what holistic perception of faces is and how we should conceptualize it. This point is a source of perpetual confusion in the research. We would like to provide a set of theoretical tools that help us understand what holistic/gestalt really means in terms of information processing properties.

Cognitive organization revealed

The structure of cognitive processing is an obscured aspect of most studies in face perception. Reports on holistic processing of faces do not usually provide testable models of cognitive organization. Rather, they are focused on problems in memory representation (also perceptual encoding), or on experimental effects that could suggest presence of configural processing which has been operationally defined. Two major categories in configural research investigation could be describe as “how” and “what”. The latter – the representational issue - is concerned with describing the object of perception in terms of its distinguishable physical properties (size, color, frequency, etc), as well as how perception of some attributes could be affected by perception of others. Such studies usually employ multidimensional scaling techniques (MDS) (Busey, 2001; Sergent, 1984; Steyvers & Busey, 2001). The interpretation of how different dimensions might interact in perception has been explored in several studies (Garner, 1976; Maddox, 1992; Shepard, 1964, 1991). However most studies are mute with respect to important aspects of configural information processing (such as *how* the processes are organized, conducted, how a decision is made and what is the capacity of the system). Several investigations are exceptions (Innes-Ker, 2003; Thomas, 2001a; Wenger & Ingvalson, 2003; Wenger & Townsend, 2000, 2001). However a study of organization of mental processes is potentially more powerful given that it enables us to look into complex processing systems, which is not possible under typical "what" studies.

The existence of some elementary, piecemeal processing representation is assumed both by the analytic approach and also by some holistic approaches (e.g. Bartlett & Searcy, 1993). We will call it face-feature, as a representational term. The idea that there is some part-face based information used in our system is backed up by findings in neural organization and specificity of our visual cortex. It was demonstrated that the analysis or decomposition of a visual scene precedes global percept, and that different neural substrates especially attuned to detect elements of visual scene are organized into more complex neural groups, which respond to more complex combinations of stimuli (e.g. Thorpe & Fabre-Thorpe, 2001).

In the next section we will review findings from several studies that suggest the existence of smaller face units, at the representational level. We will also see that different theories use features at different levels of abstraction. So, when we use the term feature we do not necessarily mean what is typically considered a face-feature: eyes, lips, nose, etc; the spatial relation between these nameable features could also be considered a feature, for example.

Face features/face-representational units

The following categorization of face-features based on the status of face representation could be identified over the body of research:

1. *Face component*, featural, componential, and piecemeal. These terms refer to smaller, visually decomposable units which are verbally namable and separable from other parts by high contrast (for example eyes), contour or texture.
2. *Relations between elements*, the configural information (spatial interrelationship of facial features). Physical face features are spatially related so many researchers provided evidence that spatial relations could be face feature.
3. The face as a *whole or holistic representation*. Raw, whole-face image is stored and then utilized later.

<u>First and second order features</u>

Rhodes (Rhodes, 1988) post a question about what the **features** of identity are? Despite the high overall similarity between faces, we are very efficient in face identification/recognition. Only a small set of cues can help us differentiate between two identical twins. Several studies questioned the importance of salience of some facial cues in face recognition (Haig, 1985; Rhodes, 1985; Shepherd, Davies, & Ellis, 1981). One way to define a face feature is the ability to label it by a single word in natural language (eyes, mouth, chin, nose, etc.). So, Rhodes (1988) defined two sets of face features involved in identification/recognition: first-order and second-order features. The *first-order features* correspond to discrete face cues that can be labeled (eyes, mouth, chin, nose, etc.). The *second-order features* are configural features, which characterize the spatial relations between first-order features and information about the face shape. Also, there exist higher-level features which are even more complex and made from the set of lower level complex features. For example age is a complex feature that contains information from both the first- and second-order features (eye size, lip thickness, etc.). The critical difference between levels of feature complexity is that the first-order features can be specified without reference to the other face features, while the second-order (configural) features must be specified in relation to some other feature. Several studies provided evidence that the higher-order features (such as sex, race, and age, emotion) are more important in face perception (Groner, 1967; Milord, 1978; Shepherd & Deregowski, 1981). Rhodes (1988) found that both the first- and second-order features are encoded and represented in the face space, as revealed by the multi dimensional scaling (MDS) solution.

Isolated and relational features

Use of the second-order configural information was further detailed by Diamond and Carey (Diamond & Carey, 1986). The authors suggested that features of the face-recognition system could lie along on a continuum which characterizes the level of reference of each feature with respect to the other features. On one end of the continuum are "isolated" features, or the features that do not make any connections to other features. This is the standard view of the cognitive system that is based on independent face features. When we move along the continuum, we increase the level of relativity of each feature with respect to other features. For example, at some point we could have color, texture of hair, shape of mustache, beard or eyeglasses. Going further on that continuum increases the relative properties these features and we get some shape face properties (a face expression, wide-set eyes, etc.). Sufficiency of the relative aspects was demonstrated in the studies were only low frequency information aided face perception (Harmon, 1973). A low frequency presented face picture bear virtually no information about features. Despite that a face could be recognized above chance. Another line of evidence comes from the studies showing that perceivers are very sensitive to tiny perturbations of internal face-spacing in photographs (Haig, 1984; Hosie, Ellis, & Haig, 1988). Diamond and Carey proposed two important face-relational properties: *first- and second-order relational properties* (Diamond & Carey, 1986). All members of visual objects that belong to the same class must share common configuration. So, utilization of the first-order properties is useful in discrimination between different classes of objects. All faces share the same configuration: eyes are above nose and lips, and nose is central and above lips. Individuation of the members of the same class is achieved by utilization of second-order relative properties such as separation of eyes, position of the lips, etc.

Face representation as a point in multidimensional space

Objects that undergo visual perception can be viewed as a point in multi dimensional visual space that serves as a meta-theoretical representation of the physical world. Correspondingly, each face could be seen as a point in a face-space. Dimensions of that space usually have some psychological interpretation, but not necessarily so (see computational methods section). Distance between face-representations is usually taken as a measure of similarity in that space, where short distances correspond to high similarity. The theoretical constructs developed in this approach are concerned with how the characterization of that space could explain a body of experimental evidence, and account for typical effects found in categorization and recognition tasks. Other aspects are concerned with the mathematical properties of the space(s) (O'Toole, Wenger, &

Townsend, 2001; Townsend, Solomon, & Smith, 2001; Townsend & Spencer-Smith, 2004), such as what type of metric should be employed, which type of geometry (spaces) should be used for characterization of faces, how to infer face-distance from the measure of their similarity, and how the representation relates to holistic or configural face encoding.

We will briefly refer to the notions of template and prototype, because they are important theoretical construct for some theories within this approach. According to *the prototype model* faces are encoded by a reference to a generalized baseline memory representation. The idea is relatively old: Galton created a prototypical face by superimposing photographs of members of the same family (Galton, 1879). The *prototype effect* is a tendency to respond to a central value of set of exemplars, even when it was not seen before (see Cabeza, Bruce, Kato, & Oda, 1999). Diamond and Carey (1986) suggested that faces share a common frame (first order relationship), and what makes them differentiable within the class of face-objects are spatial deviations from the prototypical face. We could also add that a prototype is an averaged entity of the class of objects, or the central tendency. Prototype face is located in the center of the face-space.

How does the prototype theory accounts for some findings? Similarity and distance of each perceived face with respect to the prototype plays a role in recognition and categorization. Each perceived face is compared to the prototype and the deviation from it is encoded. The stored face-information, or face features in this case, is the distance or the deviation from the prototype, and this information is utilized in face encoding and recognition.

In order to optimize recognition more prototypes could be used. For example a prototype could be very distinct from the new presented object (a new face). In order to gain more flexibility in recognition, two prototypes could be stored as a set of exemplars, defining two distinct groups or classes of objects. This approach could account for between-category findings via set of face exemplars.

Based on the role of a prototype in encoding, several approaches could be defined (Valentine, 1991). The *norm-based model* posits a face prototype in the center of the multidimensional face-space. Each new face is encoded as a deviation from the prototype as a vector in that multidimensional space. The vector points away from the central prototype, and its projection of its length onto each dimension characterizes the influence of that dimension. Similarity between faces is defined by both calculating the distance between two faces (two points in the space) and as well as the distance from each point to the origin or prototype. In contrast, in the *exemplar-based model* the similarity between two faces is solely determined by the distance between their points, while the central points play no role (Smith & Medin, 2002).

An interesting aspect of the prototype theory is that a prototype face could formed even thought the exact prototype-face has not been presented at all, simply by mental averaging over many previously learned faces (Solso & McCarthy, 1981). Inn, Waldern and Solso (Inn, Walden, & Solso, 1993) suggested that the prototype consists of the most frequent features. Bruce, Doyle, Dench and Burton (Bruce, Doyle, Dench, & Burton, 1991) provided evidence that several prototypes could be formed for several distinct groups of learned faces.

Formation of the prototype is independent of the level of information that it carries. Cabeza and Kato (Cabeza & Kato, 2000) formed two types by a morphing technique (elastical blending of pictures): featural and configural prototypes. The configural prototype was formed by morphing four different faces into one. The featural prototype emphasized the importance of face-features and was created by the morphing of features only.

Nosofsky (Nosofsky, 1991) showed that the generalized context model, which is based on exemplars, can encompass many predictions and experimental findings in face processing. This model was generalized from the Medine and Schaffer (Medin & Schaffer, 1978) the context model, which in turn, relied on the Shepard-Luce similarity choice model for identification (Luce, 1963; Shepard, 1957; Townsend & Landon, 1982). In contrast to the generalized context exemplar model the most flexible prototype model has less explanatory power. A major alternative to Nosofsky's generalized context model has been the bounded performance model (e.g. Ashby & Gott, 1988). In that model the face-space is carved up into a set of mutually exclusive and exhaustive regions, each of which is associated with a distinct response.

Gestalt and holistic properties could be captured by relatively complex metric spaces such as Riemannian metrics on infinite dimensional spaces (Townsend et al., 2001). They provided a general meta-theory that describes the position of face in an infinite dimensional space by a space function, therefore providing the meta theoretical set of tools that could be used to describe variety of manifestations of face perception.

We will return to the finite space for the moment and see in more detail how we infer some face properties using classical tools such as multidimensional scaling (MDS).

The general goal of the MDS (Kruskal, 1964a, 1964b; Shepard, 1962, 1974, 1980) approach is to find a set of points in a multidimensional space such that the distances between them are monotonically related to some observed measure of pairwise dissimilarity (non-parametric MDS). An additional aim is to find a function that relates and correctly matches the distances and the observed measures of dissimilarity (metric MDS). The procedure of finding an MDS solution is to measure pairwise similarity ratings between objects in order to find a set of dimensions and points within those dimensions that describe each compared object. The goal of the procedure is to uncover the simplest dimensional structure for the directly unobservable psychological space which determines the properties of perceptual or memory representations.

Holistic and analytic properties are examined by inspecting interactivity between dimensions. The theory behind the MDS application and the interaction of dimensions is carefully covered in the literature (Garner, 1976; Maddox, 1992; Shepard, 1964, 1991). If selective attention can be realized through the independent manipulation of each dimension then the psychological space is considered to be separable. However, interactively and holistically processed dimension are defined as being integral. Separability and intergrality are tested by comparing the goodness of fit of Minkowski power metrics. The Euclidian metric is associated with the integral dimensions, while the City-Block metric defines separable dimensions.

As Thomas (Thomas, 2001a) pointed out, several aspects of such approaches are unsupportive in understating the nature of holistic and analytic processing (by using integral and separable properties revealed by the MDS). First, and probably the most importantly is the inability to specify the exact source of possible interaction between dimensions (Dzhafarov & Colonius, 1999; Townsend & Thomas, 1993). Further, there is a problem with statistical reliability of the power parameter of the metric used (Nosofsky, 1986; Shepard, 1986).

Computational approaches

In contrast to the MDS approach, various computational approaches extract face-features from the two-dimensional image, which is analogous to the retinal image, by application of visual frequency filters that should correspond to some neural property. This approach is usually identified by utilization of the principal component analysis (PCA). Instead of frequency filters some authors use models of overlapping receptive fields (Edelman & O'Toole, 2001) or Gabor jets (Wiskott, Fellous, Kruger, & von der Malsburg, 1997). In principal component analysis, face images are projected onto eigenvectors that characterize variation of intensities of two-dimensional image at each individual pixel. All these methods provide a rich description of a face stimulus.

In computational methods, accurate holistic recognition of a face usually requires a correct localization of single face-features. For example, eigenfaces (Turk & Pentland, 1991) and Fisherfaces (Belhumeur & Kriegman, 1997) need accurate localization of key features in the face. The features could be extracted in several ways: (1) By generic methods based on edges, lines, and curves; (2) By feature template-based methods that are used to detect facial features such as eyes; (3) By structural matching methods that take into consideration geometrical constraints on the features.

Similar to behavioral-based research several computational approaches emphasize either holistic representation, featural aspects, or some combination of both. (1) In the holistic matching methods, a whole-face image is used as the raw input to the system. These methods are based on use of PCA in order to obtain (for example) eigenfaces (Craw & Cameron, 1996; Kirby & Sirovich, 1990), Fisherfaces (Belhumeur & Kriegman, 1997; Swets & Weng, 1996; Zhao, Chellappa, & Krishnaswamy, 1998) (2) In feature-based (structural) matching, local features (eyes, nose, etc.) are first extracted, and their locations and local statistics are used for structural classification. Some of prominent examples are the pure geometry method (Kanade, 1973), the dynamic link architecture (Okada et al, 1998)(Okada et al., 1998), the hidden Markov model (Nefian & Hayes, 1998). Finally (3) hybrid methods use both local features and the whole face region in recognition. Several approaches can be distinguished: the modular eigenfaces (Pentland, Moghaddam, & Starner, 1994), and the face region and components (Huang, Heisele, & Blanz, 2003).

Computational methods: what are the features?

The *cognitive* models of face perception (Cottrell, Dailey, Padgett, & Adolphs, 2001) are between psychologically plausible models and the engineering-based approaches, and should provide a basis for the induction leap, concerning the characterization of face processing system. A benefit of that leap depends on (a) the biological plausibility of the model, (b) the extent to which actual model performs the same task as humans, (c) the correlation between human measures and the face model predictions, (d) the extent to which the model provides useful insight into the nature of the process of face perception and (e) novel predictions that the model can account for.

Cottrell (Cottrell et al., 2001) defines several dimensions of feature-space used to characterize features that can be extracted and manipulated using computation methods. Rhodes (Rhodes, 1988) noted a difference between two types of features that could be extracted from an image: the first-order features (simple face-features) and second-order features that are a combination of the former ones. Therefore, there is very close similarity between these two approaches based on the type of face representations used, though the differences stem from the different assumptions concerning the level of characterization of the mechanisms involved in face perception, as well as the processing mechanism. Eventually, after enough research evidence is accumulated, both approaches should converge and reveal the same cognitive structure. At this moment the computational approach makes more bold assumptions about how the extraction process is conducted by freely manipulating several representation properties. According to Cottrell (Cottrell et al., 2001) the dimensions of face representations are (1) local/global depending on spatial context of the features relative to the object of interest (features of holistic images are usually treated as memory templates of different complexity level); (2) the rigidity of a particular region of interest, which specifies to what extent the feature or templates can move on the face image, and to be deformed in order to account for the variability of their appearance on the face; and (3) whether the features are learned from the examples in the domain of interest.

Cottrel (2001) differentiated between three type of features derived from raw two-dimensional face images using PCA: the eigenface, eigenfeature and local principal components. The *eigenface* is an eigenevector of the covariance matrix of the face image set. The *eigenfeature* uses the same extraction process except that the analysis is restricted to the rectangular regions around the eyes and mouth (for example). The l*ocal principal component eigenvectors*: provide a "basis image" and it resembles the filtering performed by cells in primary visual cortex. Computations are conducted on a large number of small pixel patches sampled uniformly from random positions on the face.

The general criticism of the computational approaches is the lack of psychological plausibility of the theoretical constructs and the fact that there is no clear and strong relationship between behavior and properties of these models and observed human performance in various cognitive tasks.

Processing models

Face-space and processing structure

In previous section we stated that the two most fundamental questions in face perception (and in object perception as well) are concerned with providing a detailed status of face representation and organization of mental processes during face perception. The two issues are closely interlinked, and some recent attempts have been made in order to specify this connection (Sergent, 1984; Thomas, 2001a; Townsend & Thomas, 1993; Wenger & Ingvalson, 2003; Wenger & Townsend, 2001). To some extent both the "who" and "what" approaches seem to be complementary. By examining the face-space and providing a reasonable approximation of its dimensions, we get rich information of what could be the building blocks of face representations, which are connected to configural/holistic properties of a face representation. If we can validate that particular face representations are holistically memorized, then the face-space describing it should be different, in terms of the metric used, and the properties and relations between dimensions, when that holistic face representation is compared with some other representation that does not posses holistic properties. In fact in order to show that face representations are based on analytic mode of processing, which is analogous to the sum of independent face-features, a MDS solution of the similarity of an orthogonally manipulated set of faces should reveal an orthogonal solution with the same number of dimensions. Several seminal studies for both separable and integral dimension stimuli were conducted with an identification-categorization task (Nosofsky, 1986, 1987; Shepard & Chang, 1963; Shepard, Hovland, & Jenkins, 1961).

Usually that approach suffers from being relatively static when describing perception: what happens in the cognitive system in a more general case when learning occurs, and how we actually map the point-wise face representation in that space remains more or less imprecisely defined and unclear. Here we care about the processing structure during face perception. We will provide one representative example which clearly warns the modelers when bypassing the question concerning architecture of face-encoding. By the architecture we mean whether processing is organized in serial or parallel, in the simplest form possible. So, when a face representation is mapped onto some face-space, by application of MDS for example, we really do not know how the cognitive system used the dimensions of that space. We do not know whether they are accessed in serial or in parallel, for example. Because dimensions can be seen as sources of evidence, and both serial and parallel architectures could collect evidence from orthogonal, i.e.

independent, dimensions. Even when parallel architecture is limited capacity, independence, or orthogonality in this case, can be preserved, and both architectures could provide the same MDS solution. Some MDS modelers stick with the parallel processing architecture (e.g. Nosofsky & Palmeri, 1997) and some with the serial (Tversky & Krantz, 1969). Seriality in the latter case is based on an unsupported assumption that orthogonality is equal to analytic processing, which is considered to equate with serial processing. A theoretical issue that is closely related to that problem is model mimicry (Townsend, 1971a) between serial and parallel structures, and is especially hard to solve when simple mean response time analyses are used.

In addition to the issues with architecture, the MDS solution does not indicate at what level there was interaction between dimensions of interest. Usually, the dependency between dimensions of face space is solved by determination of the space metric properties (separable or non-separable dimensions). But possible levels of interaction are numerous: information could interact at stimulus input, during evidence accumulation, at the processing output, or at the decisional stage (Ingvalson & Wenger, 2005; Thomas, 2001a; Wenger & Ingvalson, 2003).

We will also see that sometimes specific treatment of the face-representation dictates the structural issue, i.e. the format of a face representation dictates how processes are organized. The two issues of representation and structural organization could be tightly interwoven

Typically, there is a tendency to equate or make functionally inseparable the two issues. So for example, analytic processing which assumes piecemeal, feature-like representations is usually defined as a slow serial process (e.g. Tversky & Krantz, 1969).

In contrast, holistic processing (or perception), which assumes template like representations, calls for parallel processing.

The advantage of using more powerful modeling is the ability to test the system's architecture, dependency, stopping rule, and capacity. We will suggest stronger tests for processing structure in a later section. Meanwhile, in next section, we will focus on the history of developing the ides of face processing models.

Serial processing

In the experiment of Smith and Nilsen (Smith & Nielsen, 1970), participants had to make same/different decisions between two faces, while number of dissimilar features was manipulated. The time needed to make response decreased as the number of dissimilar features increased. They found that 'same' and 'different' responses are differently affected by the time delay between faces that were compared. 'Different' responses showed more sensitivity to additional differing of feature-dimensions. 'Same' responses showed sensitivity to the addition of more relevant dimensions to the faces, only at lag of 10s. The authors suggested that a feature comparison process occurs on 'different' trials, while 'same' faces are processed more holistically. In fact, with prolongation of the delay between the two faces, the observed holistic processing of the 'same' response switched to a more featural, independent processing strategy. So, the main assumption concerning identifiabilty of analytic and holistic processing is sensitivity (at the mean RT level) to increasing the number of relevant dimension. If mean RT is not affected by this increase, then a holistic processing structure is suggested. By inspection of mean reaction time for the presence of a particular feature, they also suggested that faces are scanned from the top to the bottom in the serial-like manner. This finding was also consistent with analytic processing. When to-be-compared faces were separated by 1s during presentation, the manipulation of the number of similar features did not show any effect in the "same" condition. That was taken as evidence for a *template matching process*. Thus, they adopted a dual-model, for face recognition: both analytic (serial) and holistic (template matching) could be used, depending on properties of the faces to be compared. A possible confouding factor was revealed subsequently: the number of relevant features is correlated with the number of irrelevant features since the faces used in experiments had five features (see Sergent, 1984).

Further support for serial processing model came from Bradshaw and Wallace (Bradshaw & Wallace, 1971), and Hole (Hole, 1994) who suggested that a "same" decision takes more time than a 'different' due to the engagement of serial, self-terminating search. Serial self-terminating search means that participants checked one face feature at a time: say nose, then the eyes, then the hair, until they found any difference. If no differences were observed then they responded 'same'. Further, Rhodes (1988) suggested that faces are processed serially top-to-bottom, emphasizing the particular importance of eyes in identity recognition.

The strict serial architecture is usually enforced by the idea that there exists a feature processing order. As indicated in Smith and Nielsen (Smith & Nielsen, 1970) the order of processing could be revealed by mean reaction time for particular component. If the standard serial system is employed during face perception then processing order could be determined by the difference in reaction time needed for each component to be recognized. However, we know that the limited capacity parallel processing model can mimic behavior of the standard serial model (Townsend, 1971a), and the processing order in the parallel system is determined probabilistically by the ratio of processing rates between channels. If processing rates for different face-features are different, then mean reaction time for recognition for each component could mimic behavior of standard serial system. Therefore, more powerful tests are needed for exact determination of the system's architecture, and such measures could include inspection of termination rule as well as and the capacity (Townsend & Ashby, 1983; Townsend & Nozawa, 1995)

Although the fact that the difference in mean recognition RTs between face features cannot be used to reveal architectural properties, it could be used as evidence against a strong version of the template model. The template model assumes that faces are encoded into icon-like, and unitary face representations, where extraction of face parts, or face features, requires approximately the same amount of time. Given that different feature-sensitivity was observed in many investigations (Bruce, Dench, & Burton, 1993; Davies, Ellis, & Shepherd, 1977; Ellis, 1975; Rakover & Teucher, 1997) it could be argued that features, do exist and that they are encoded as independent representations (e.g., Sergent, 1984).

The dual-mode hypothesis, which assumes that the cognitive system relies on both featural and holistic sources of information was proposed very early in face processing research (Smith & Nielsen, 1970). It is interesting to note that the dual status of face representations is usually automatically connected to the different processing systems that include serial and parallel processing. We will discuss these issues in the next section.

Top-to-bottom face processing has also been evidenced in the work of Sergent (Sergent, 1982). It is generally found that top facial features provide more information than bottom ones. In the study of Shepherd et al (Shepherd et al., 1981; p.105) it was found that the participants could rank the face-features with respect to how long they held their attention to catch: eyes (62%), hair (22%), mouth (8%).

Further evidence for analytic processing was suggested by Walker-Smith (1978) and also by Tversky and Krantz (Tversky & Krantz, 1969) by showing that the contribution of each facial component was independent of other face-components, which the authors described by using serial processing model.

<u>The parallel processing models</u>

Matthews (Matthews, 1978) was the first to suggest that perception of faces includes a mixture of parallel and serial processing. He used the 'same-different' method as well, and found that changes to either the hair, eyes or chin are detected equally fast, and these features are faster then the eyebrows, nose and mouth. However, this conclusion could be questioned (see critique by Sergent, 1984). Averaging across different participants could eliminate or mask individual differences that might exhibit a strictly serial processing strategy. In fact, Sergent (1984) suggested the parallel processing architecture with dependent processes. Sergent pointed out a very important general problem with the same/different reaction time approach: if we vary a number of different features, then we ultimately change configural properties as well. If a serial strategy is likely to be adopted, then it is not certain whether this is because of the change to a number of different features, or because there is a change in configuration. A recent publication (Wenger & Townsend, 2001) revealed support for parallel self-terminating face processing, using stronger tests for architecture based on the system factorial technology (SFT).

The dependency issue

The issues concerning processing architecture in subsequent investigations are practically bypassed, and other means of modeling are used in order to explain the data. The reason for avoiding analysis of architecture was twofold: the mimicking between serial and parallel models renders some architecture indistinguishable when means are used (Townsend, 1971b); so without a strong test not based solely on means it seems much easier to adopt the difference feature- and holistic-based representations as an explanatory tool. The feature representation invoked the so-called analytic processing mode, and everything else was termed the holistic/configural processing of faces, although in many variations. This combination of both holistic and analytic modes provided a very flexible alchemistic combination that was very powerful in data explanation, although complicated to falsify.

Instead of further developing and testing the validity of some processing models (based primarily on architecture) most research has adopted the assumption that governs simultaneous usage of both (or many) formats of face-representation during face encoding. The pitfalls of avoiding architectural issues were obvious; for example, Sergent (1984b) found that the decision time to detect two features is shorter then the time to detect only one. In the same/different task she manipulated three dimensions: eyes, chin contour, and "internal space" (whether the eyes and nose were closer to the forehead or the chin) and demonstrated decrease in mean reaction time as she increased the number of "different" dimensions. Note that only one face-feature (dimension) is needed to make a correct "different" response. She found that when she increased the number of different features then the decision is faster. This was considered clear evidence for a facial-representation that consisted of dependent face-parts.

Sergent (1984) aimed to investigate dependency between facial features rather than focusing on the architectural issues. But this is *prima facie* evidence of how knowledge of processing structure can help explain findings. If face processing is independent, that is if the processing time of each face-feature is by not affected by processing time of other face-feature (stochastic independence) regardless of architecture, and can terminate on completion of any feature (self-terminating or minimum time processing in this case), then determining whether faces are same/different will depend probabilistically on the speed of processing of the most salient or fastest feature.

It is well known that the presence of two signals, which are both positive targets, will produce faster RT when compared with the single condition, even when the same processing rate parameters are used in both conditions (e.g. Colonius, 1990; Grice, Canham, & Boroughs, 1984). Such statistical facilitation occurs with any number of multiple targets. The class of models that can produce statistical advantage in the redundant target case is the minimum time, independent, parallel processing system (the so-called horse race model). Thus we have an example of a particular parallel independent architecture that can be confused with the model that assumes dependency, as in Sergent (1984).

Configural and Holistic processing

Mondloch, Le Grand and Maurer (Mondloch, Le Grand, & Maurer, 2002) broadly defined *configural processing* as perception that involves perceiving relations between the features of a stimulus. So according to the authors, configural processing could be divided into three types: (1) sensitivity to first-order relations, that is, recognizing a face because the eyes are above the nose, and both are above the lips (Diamond & Carey, 1986; Johnson, Dziurawiec, Ellis, & Morton, 1991; Kanwisher, Tong, & Nakayama, 1998; Moscovitch et al., 1997); (2) holistic processing, which is based on gluing together the features and forming a gestalt (Hole, George, & Dunsmore, 1999; Young, Hellawell, & Hay, 1987) – higher recognition accuracy is achieved when a face-feature is presented in the context of previously learnt features (Tanaka & Farah, 1993; Tanaka & Sengco, 1997); and (3) sensitivity to second-order relations, which are defined as information concerning the spatial relations among features (Leder & Bruce, 1998, 2000). According to the authors, the configural is a more general term than holistic, since the latter is a special case of the former.

For Gauthier and Tarr (Gauthier & Tarr, 2002) the term holistic is a super-ordinate term and cannot be defined using a single mechanism or representational format. Based on separability of experimental findings, they operationally defined holistic by: (1) Holistic-configural, referring to the experimental finding showing unique effect of configural processing on part identification (e.g. Tanaka & Sengco, 1997); (2) Holistic-inclusive, referring on obligatory processing of all features from an object compared to when a feature of interest is combined with different parts of another object; and (3) Holistic-contextual, arising when individual parts are better recognized in the context of other parts then in isolation (e.g. Tanaka & Farah, 1993), and is independent of expertise, unlike the first two. Bartlett, Searcy, and Abdi defined **configural processing** as a broader category than **holistic** processing given that configural need not be holistic (Bartlett, Searcy, & Abdi, 2003). **Configural** processing (a) involves template-like structure (we will call it the stronger assumption of configural processing) that probably includes the whole face (Bartlett & Searcy, 1993; Farah et al., 1998; Yuille, 1991), and (b) it is sensitive to the internal facial region, as was revealed in a patient that suffers from object agnosia, but not prosopagnosia (Moscovitch et al., 1997), and (c) spatial relations between adjacent features are encoded as local representations (such as distance between the eyes, mouth and tip of nose, etc.), rather than being global or holistic.

We will focus now on several approaches that are more explicit in terms of their definitions, but they are less general as well. Two definitions of holistic processing have been offered by Farah, Tanaka and Drain (Farah et al., 1998) and Carey and Diamond (Diamond & Carey, 1986): **holistic encoding** and **configural encoding**.

In **holistic encoding** parts of perceived objects are not represented and/or processed independently but are perceived as a whole. That definition is very close to the idea of a template matching scheme. A face representation does not consist of parts or face features, but rather the features form a unitary representation. Accessibility of the parts is different than the whole face.

It is debatable whether the unitary form of a holistic representation should always help part processing. That is, will a template face structure help or hinder recognition of a face feature? Some convention is that good form or gestalt should hinder processing of a part (Donnelly & Davidoff, 1999; Tanaka & Farah, 1993; Tanaka & Sengco, 1997).

Experimental evidence suggests this is so: a holistic representation can inhibit the search for an individual feature .

The second definition, **configural encoding**, refers to the idea that the spatial relationship between features in a perceived object is a major determinant in holistic perception. A representation of an object consists of separate feature representations. Also, spatial relations between these features are stored, and they provide a crucial contribution to holistic processing. Tanaka and colleagues (Tanaka & Sengco, 1997) provided a demonstration of how single feature is more accurately recognized when presented in a face background than in isolation.

Holistic vs Configural vs Featural

Sometimes the terms *holistic* and *configural* are used interchangeably in the literature, or sometimes they reverse generality with respect to each other: for example, according to some authors (Bartlett et al., 2003; Mondloch et al., 2002) configural is broader then holistic, while for others (Gauthier & Tarr, 2002) the reverse is true. It is not surprising than that they are frequently confused.

Part of the confusion stems from the fact that no absolutely clear definition exists of what the individual face features are. Processing face features should be on the opposite side of dimension of what we define as a holistic/configural processing. Let us see one useful categorization of face-features that is derived from the holistic definition (Bartlett et al., 2003): (1) face-features should be explicitly represented in memory codes; (2) they are consciously accessible for verbal report; (3) they are encoded in such way that no other feature representation influences them. We could agree with this definition, although we will consider some problems.

Features could also have a configuration; for example, in the definition of the first- and second-order relational properties (Diamond & Carey, 1986), face-features on the level of cognitive representations are encoded as spatial relations, not particular features per se. We can also think of recognizable face features as possessing holistic properties: nothing prevents us for treating an eye as a gestalt. It has some form; it is comprised of separate feature, and we also have appropriate verbal names for the features that constitute an eye (iris, cornea, pupil, eyelid, lashes, etc.).

So apparent confusion of terms configural and holistic has reasonable grounds: definitions of parts and their relations can be arbitrary. Solutions for the feature definition could be to treat the smallest part of face as a feature (for example, as extracted in the computational approach), or we can look for face parts that are visually discrete and can be separately labeled. But then the definition of holistic/configural is troubled - there are infinitely many possible spatial relations between infinitesimal face parts, and therefore infinitely many levels that holistic/configural properties could be defined. In fact, a face and its parts are a complex, spatially-nested structure that possesses a good form and can always be broken into smaller feature units. It is almost arbitrary to define the cut-off point, to define the features and to define the relations between those defined features.

Since the definition of holistic/configural processing suffers from being an impenetrable construct, both the terms holistic and configural are usually operationally defined in experiments that produce similar experimental effects. The good news is that recently we have seen some consensus concerning the definitions of holistic/configural.

However, in order to properly define the terms we need to postulate either the **mechanism (processes)** of that processing and/or **representational status** of face-parts. The last thing we want is to define these terms based on operational definitions within the experimental paradigms.

The goal of many researchers in the face recognition field is to provide an accurate definition of holistic face perception. A definition through operationalization is one way to do this, but in this case it has produced lot of the confusion and disagreement. Again, this is mainly because theoretical constructs are uniquely identified with the experimental findings, which is the operational definition procedure by itself. Some cautious should be raised here, and we must suggest that an operational definition should and must be connected with the theoretical validity and uniqueness of the construct. The pitfalls of simply naming the experimental effects without thorough theoretical investigation including both representational issues and processing structure are unavoidable. We recall debatable Sergent (1984) finding of feature dependency mentioned earlier that could be explained twofold, but this is not isolated case.

We suggest that in providing a general theory of (special) face perception, more attention should be devoted to testing the validity of theoretical constructs. If it is possible that the same experimental effects can be accounted for by different organizational principles, then some operationalizations are seriously questionable.

The question of the building blocks of the face representational system is not only what the smallest parts (atoms) are, but also, what are the relationships between them, and what imposed structure of relations governs their utilization. Do featural (part-based) face representations with different levels of complexity exist and what is their relationship to each other, in terms of a relationship between holistic and analytic processing? We suggest that in the cognitive system, all faces are represented with their smallest units operating independently. If the smallest units are namable face features (eyes, lips, etc.) then several independent mechanisms could collect the evidence for each feature presence. Another system is needed to integrate the information coming from all these independent detectors in order to make a decision. This is a simple description of *the independent feature processing model*. This is probably an oversimplified view of our face cognitive system, but it can serve as a comparison model.

Given that we have a reasonable abstraction of what the features are, then the questions of interest are: what are their relationships to each other, and can they be represented at different levels of abstraction? The former question regards interdependence between representational features (testable with general recognition theory, GRT), and the latter concerns between-feature relationships. For example, simpler units can make more complex ones (lips and chin could combine into a lower half of the face feature). Does that more complex unit exist on its own, or is it constructed from the lower units for each percept?

In the next section we will outline major research advances in differentiating between the statuses of the feature representations. The approaches covered here are the feature based dual-mode hypothesis, interactive hypothesis and holistic processing theories. The two issues of concern here are the status of feature representations and the relationships between them.

Featural hypothesis (feature-based)

The discrete facial features (eyes, nose, lips, etc.) are very important in many human activities, including many aspects of social life and communication. Some features are particularly important in conveying verbal information such as the lips; another example, the eyes make the difference between true and false laughter (Ekman, Friesen, & Ellsworth, 1972). Face features provide cues to understanding someone's mood, gender, and age. It seems, then, that the featural information, given its importance in conveying information, is very important - probably more than configural information (Rakover & Teucher, 1997).

One of the earliest pieces of strong evidence for independence of face features was demonstrated by Tversky and Krantz (Tversky & Krantz, 1969). They applied an MDS analysis of dissimilarity judgments between pairs of faces. They demonstrated that face features are independent and that the whole face experienced as the sum of perceptual effects coming from each feature. These findings were criticized on the basis of using schematic faces which appear non-realistic.

In contrast, in face recognition studies when face-encoding instructions are manipulated then the faces are recognized on a configural rather than analytical (featural) basis. Recognition is more accurate when it follows global evaluative judgments (for example, likeability, attractiveness, honesty, personality) than when judgments are focused on isolated features (Bower & Karlin, 1974; Patterson & Baddeley, 1977; Winograd, 1976).

Independent configural and featural (Dual-mode hypothesis)

The idea of the dual-mode hypothesis is that both featural and configural modes of processing operate simultaneously and independently (Bartlett & Searcy, 1993; Bartlett et al., 2003; Searcy & Bartlett, 1996). Separability between configural and featural processing is evidenced in multiple studies. Some evidence comes from different sensitivity to face manipulations such as face inversions (Bartlett et al., 2003). The face inversion manipulation is believed to be the most ubiquitous factor that differently affects whole faces and face part stimuli. Recognition of features is usually affected by inversion to a smaller extent than whole faces. Additional evidence comes from a different role for learning and retention of faces. Reinitz et al. (Reinitz, Morrissey, & Demb, 1994) suggested that holistic representations are not directly stored in long-term memory, but are reconstructed at retrieval based on memory for features and the relationships between them. Bartlet (Bartlett, 1993; see also Bartlett et al., 2003) found that in a face recognition memory task, a new face composed of previously viewed features is often recognized as old. Although this finding could be attributed to featural processing only, converging evidence from other behavioral findings, computer simulation and neuro-imaging studies revealed influence from both holistic and featural processing. Finally, featural and configural processing are anatomically distinct (Rhodes, 1985, 1993) and show different patterns of hemispheric asymmetries. It has been demonstrated that right hemisphere activation extends more anteriorly to include the fusiform face area (FFA) (Rossion et al., 2000).

Although there is converging evidence of independence between featural and configural modes, stricter and stronger tests are needed. For example, providing significant effects for the contribution of both models in a behavioral study does not discount possible interactions and interdependence. Similarly, evidence of separability of neural structures does not prove the non-existence of an interaction. For example, Rakover and Teucher demonstrated that recognition of features outside face context is affected by face inversion. Thus, features are affected also but to lesser extent than whole faces (Rakover & Teucher, 1997).

Stronger test of reliance on features only during face perception is evidenced in Macho and Leder (Macho & Leder, 1998). They pointed out weaknesses of previous investigations and the importance of scaling issues in the dimensions of interest. Using the logit model, which is more suitable for analysis of proportions and allows for hierarchical testing of complex interactions between featural and configural (holistic) sources of information, they found that, in face processing, the main contribution come from the featural aspects. In their study, they manipulated the size of the face features (lips and nose) and the configural aspect (eye separation). No interaction was observed between pairwise tests of each feature and the configural aspect. Therefore they are assumed to exist as independent face properties. Besides its exact statistical appropriateness in investigations dealing with proportions, application of the logit test possesses lower statistical power than the general linear model (GLM) which is not the appropriate test for proportions or errors. However, use of the GLM is suggested in reaction times studies given that reaction time lie on a ratio scale, and it provides a more statistically powerful test.

Leder and Bruce (Leder & Bruce, 2000) found that featural information did not depend on the configural information. In one of the conditions, participants had to recognize a face feature, presented in isolation, compared to the case when the same feature was presented together with the addition of a redundant face context (in order to encourage holistic processing). They suggested that manipulated relational features were locally encoded and probably not *post-hoc* encoded from a face-template representation. A similar finding was provided by Macho and Leder (Macho & Leder, 1998) who demonstrated that relational information (eye separation) did not interact with the availability of other local features, in a face similarity decision task.

Cabeza and Kato (Cabeza & Kato, 2000) showed that both featural and configural information are encoded and stored. They formed featural and configural prototypes and demonstrated that both are sensitive to featural and configural manipulations. The participants showed a tendency to commit a false alarm when they were presented with a previously unseen prototype, and this tendency was equal for both prototypes. Further, separability between the two sources of information was suggested when the two prototypes showed different sensitivity to face inversion.

<u>Interactive configural and featural</u>

Young, Hellawel and Hay (Young et al., 1987) demonstrated that configural and featural aspects can interact (see also Hole, 1994). They misaligned the upper and lower halves of several faces and asked participants to recognize either the upper or lower half. The aligned combination of old and new features hindered participants' ability to correctly recognize single halves. Horizontal dislocation of the halves improved correct identity recognition of halves. Combining halves from different faces prevented holistic encoding of single halves, probably by interference between featural and configural information preserved in both halves.

The interaction between the face features and the perception of emotion has been demonstrated as well in face expression research. Different facial regions interact to produce different emotional expressions (Ekman et al., 1972). For example, the same movement of the brow may convey different emotions depending on the movement of the mouth (Ekman & Oster, 1979; McKelvie, 1973).

We will also review the holistic encoding hypothesis (Farah, Tanaka, & Drain, 1995; Farah et al., 1998; Tanaka & Farah, 1991, 1993; Tanaka & Sengco, 1997) that also includes both featural and configural modes. The **strong version** hypothesized that faces are encoded and stored in a way that involves very little featural decomposition. There is a strong reliance on configural properties.

The **weaker version** assumes that both featural and configural aspects of face-representations contribute to face perception and recognition (Tanaka & Sengco, 1997), but configural information is disproportionaly more utilized than featural. According to Tanaka and Sengco (1997): *Holistic* approach assumes that *featural* and *configural* interact in the face representation. They are stored together and interact. Change in one type of information will change the other one. The exact characterization of a mechanism of the interaction was not offered.

Application of the strong test for processing organization and independence between processing modes (Ingvalson & Wenger, 2005) unequivocally revealed that the interactive (non-independent) relationship between configural and featural modes exist, and that they operate simultaneously. That is very strong support for the work of members of the Farah and Tanaka group.

Holistic processing

Even when images are extremely blurry it is possible to recognize faces (Harmon, 1973). Single features do not contribute to recognition at this low resolution. Ginsburg (Ginsburg, 1978) suggested that the low-spatial frequency information (at about two cycles per degree of visual angle) carries the basic face information. Higher levels of resolution are needed in order to perceive the featural information.

The strongest holistic level hypothesis assumes that faces are encoded as the templates. The template preserves a detailed face description, almost in a raw image format, and is considered to be an exact copy of a perceived face. Individual face features can not be selectively accessed from it. This strongest assumption level has its disadvantages: the rigid template representation should pose no difference on performance when different isolated face features are searched (Sergent, 1984). Also it is cumbersome for any system to deal with the number of templates needed to describe more realistic faces (or objects) that are perceived from different angles. The idea of template representations resurrected when at least some rigid assumptions were relaxed. For example, if we assume that some transformation operates on a template by means of some scale transformations and 3-D rotation (e.g. Tarr, 1995) it is possible to reduce the need for large storage size needed for the template model. Also spatial transformations could accompany several stored templates (e.g. Ullman, 1991) (for review in object perception, see Palmeri & Gauthier, 2004). In the case of face perception we could assume that the system stores different types of templates, including both featural and configural information. Thus, the face recognition system can achieve flexibility and generalizibility (e.g. the prototype model of Cabeza & Kato, 2000).

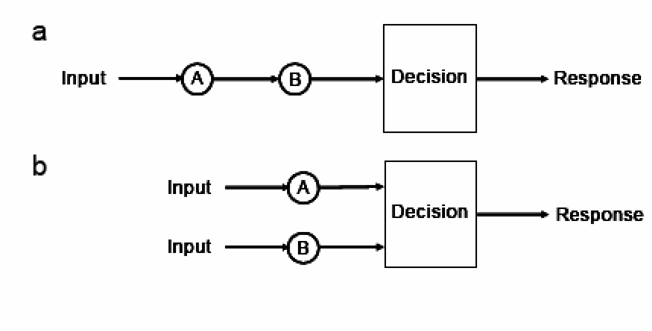
We could adopt the convention that holistic or a whole in face perception is a hierarchically higher-order term that spans across a large part of a face (Bartlett et al., 2003), encompassing lower-order terms such as the encoded features. The features must be defined on a basis of testable and observable properties with converging of both the statistical evidence of their perceptual significance and the configural information that is defined as a functional and measurable relationship between supposed features.

Defining configurality in terms of processing characteristics: Systems factorial technology (SFT)

Our approach to defining configurality looks at the issue from the perspective of four general characteristics of real-time information processing. These characteristics can be used to describe *any* information processing system that is operating on more than one encoded source of information. Analyzing and formalizing a definition of configurality in this way allows us to take advantage of almost three decades worth of important theoretical and methodological advances, (Schweickert, 1978; Schweickert, 1983; Schweickert, Giorgini, & Dzhafarov, 2000; Schweickert & Townsend, 1989; Townsend, 1972; Townsend & Ashby, 1978; Townsend & Nozawa, 1995; Townsend & Schweickert, 1989) advances that allow for strong-inference tests of a complete set of hypotheses regarding these characteristics. We refer to this body of work as *systems factorial theory* (SFT).

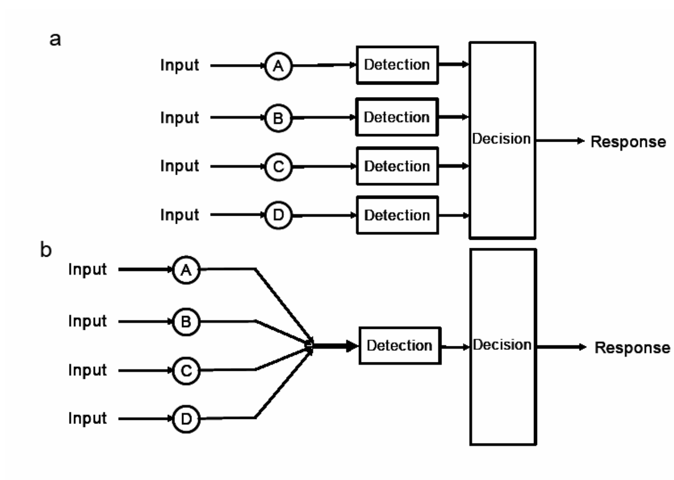
We begin with the issue of *process architecture*. Generally, there are three theoretical alternatives to consider. First, one feature may be processed first, followed by another feature. In this case, processing would be characterized as reflecting a serial architecture (Figure 1a). Second, both features may be processed at the same time; in this case, processing would be characterized as reflecting a parallel architecture (Figure 1b).

Figure 1: (a) A serial system and (b) a parallel system. The input is a source of information for the system, for example a face or a non-face stimulus. "A" and "B" denote two channels of processing, two processes, or two units. For example "A" and "B" could be face-feature detectors (responding to the presence of an eyes and lips). In a serial system both channels process the input information in a non-overlapping manner, while in a parallel system the channels operate simultaneously. After all channels finish processing (for example, the recognition of a face feature) the decision is generated. In other words, upon the positive recognition of all face features the response "I see a face" is generated. Otherwise the response "This is not a face" is generated.



Finally, both features may be initially processed in parallel, and the outputs of this processing may be pooled or combined into a single "channel" of information. In this case, processing would be characterized as a special form of a parallel processing architecture known as coactive processing (Figure 2b) (Diederich, 1995; Miller, 1982, 1991; Mordkoff & Egeth, 1993; Mordkoff & Yantis, 1991; Townsend & Nozawa, 1995). The notion of configural processing, which suggests simultaneous use of all sources of information, would seem to be at odds with a serial architecture. This suggests that configurality should be associated with a parallel or coactive processing architecture.

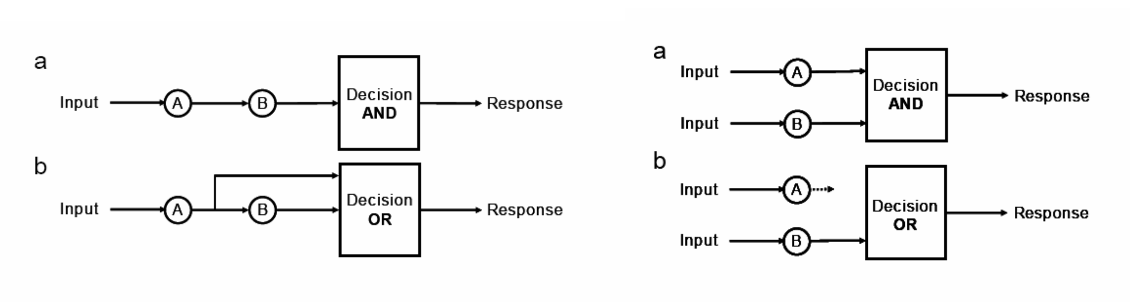
Figure 2: Schematics of (a) A parallel independent system and (b) A coactive multiple channels processing system. A coactive model assumes that an input from separate parallel channels is consolidated into a resultant common processor, before a decision is made.



The issue of *stopping rule* refers to the amount of information that is required in order for an observer to select a response. Two alternatives are generally considered. On the one hand, it may be that once some sufficient amount of either feature has accumulated, a response can be made. In this case, the stopping rule would be referred to as a self-terminating or minimum-time rule (Figure 3b). On the other hand, it may be that there is some requisite amount of both features required before a response can be made. In this case, the stopping rule would be referred to as an exhaustive or maximum-time rule (Figure 3a). The notion of configurality would seem to be at odds with the use of only some minimum amount of information. This suggests that configurality be associated with exhaustive processing.

Figure 3: (**A**) Schematics of stopping rules in a serial system. (a) A diagram of the standard serial system in the case of AND (exhaustive) processing. (b) The stopping rule in the serial system is depicted as an additional arrow which goes from the output of "A" directly to the decision box, allowing for the possibility of bypassing process "B". When the evidence accumulated by process "A" is enough to make a decision then the processing can terminate, and additional processing of "B" is unnecessary.
(**B**) Schematics of stopping rules in a parallel system. (a) A diagram of the standard parallel system in the case of AND (exhaustive) processing. (b) In the OR case, the evidence accumulated by process "B" is enough to make a decision and processing can terminate, even "A" has not finished yet processing (this is indicated by the short arrow).

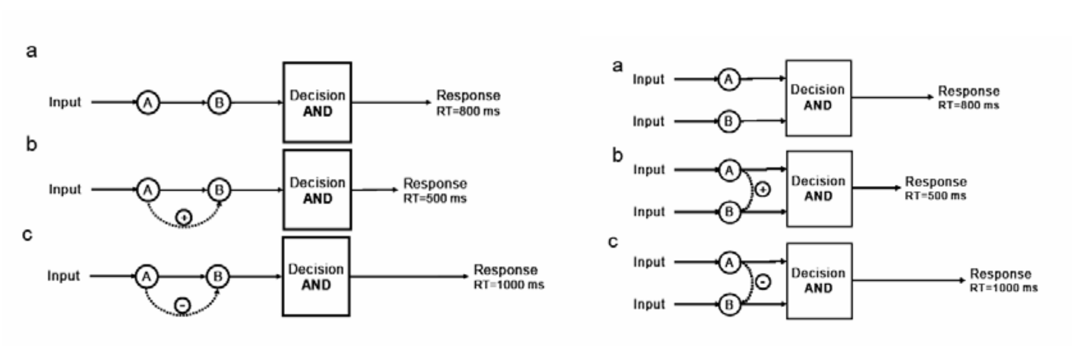
A **B**



The issue of *independence* (Figure 4) refers to the effect that the processing of one type of information may or may not have on the rate at which another type of information is processed. The first of the two alternatives needing to be considered is that processing featural information has no effect on the *rate* at which configural information is processed, and vice versa. In this case, the rates of processing of the two sources of information would be characterized as being independent. In contrast, it might be that one feature might increase the rate at which another feature is processed, or it might decrease the rate at which second feature is processed. In both of these cases, the rates of processing would be characterized as being either positively or negatively dependent. The notion of configurality would seem to be at odds with independence, suggesting that configurality should be associated with dependencies in rates of processing.

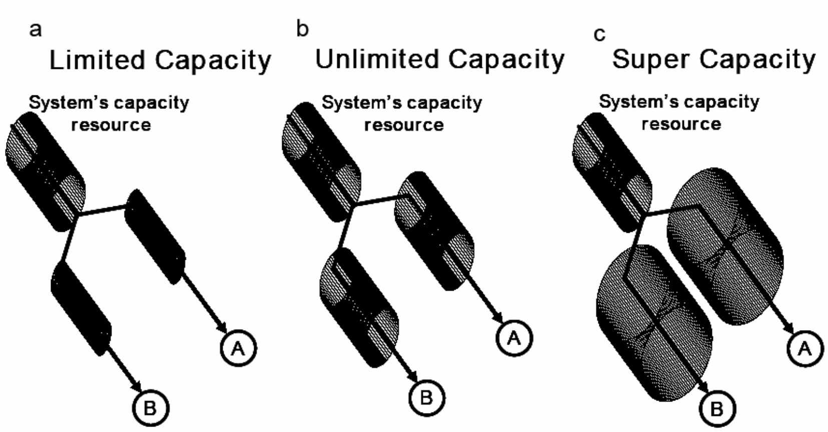
Figure 4: (**I**) Dependency between "A" and "B" in a serial system. (a) The standard serial independent system, (b) A positively dependent serial system: duration of "B" depends positively on duration of "A", that is, faster processing in "A" will produce facilitation or faster processing in "B" and vice versa. For example, in a face recognition task faster recognition of the first face feature could give some "confidence" to a second process to speed up processing of a second feature. (c) A negatively dependent serial system: the processing time of "B" is inversely related to a processing time of "A". Faster processing of "A" produces slower processing of "B"; that is, "A" inhibits "B", and vice versa. Overall, a positively dependent system with the facilitation exhibits the fastest reaction time (500ms), while a negatively dependent system with the inhibition exhibits the slowest reaction time (1000ms).
(**II**) Dependency between "A" and "B" in a parallel system. (a) The standard independent parallel system, (b) A positively dependent parallel system: The positive arrow from "A" to "B" indicates positive facilitation. That is, faster processing of one channel speeds up processing in the other channel (as depicted in the figure), and vice versa. (c) A negatively dependent parallel system: the processing time of "A" is inversely related to the processing time of "B". Faster processing of "A" will produce longer processing of "B"; that is, "A" inhibits "B" (as depicted in the figure), and vice versa. Overall, a positively dependent system with the facilitation exhibits the fastest reaction time (500ms), while a negatively dependent system with the inhibition exhibits the slowest reaction time (1000ms).

I **II**



The issue of *capacity* (Figure 5) is closely related to the issue of (Townsend & Wenger, 2004), and refers to the way in which system performance changes as workload is varied (Townsend & Ashby, 1978; Townsend & Nozawa, 1995; Townsend & Wenger, 2004; Wenger & Townsend, 2000). For example, if it is assumed that when one feature information is augmented by another feature information (i.e., if the amount of information to be processed, and thus the workload, is increased), there will be no observable effect on processing efficiency, then the perceptual system can be assumed to possess unlimited capacity (Figure 1b). If, in contrast, it is assumed that augmenting one of these sources of information with the other will result in a decrement in performance, then the assumption is one of limited capacity processing (Figure 5a). Finally, if this same change in workload produces an improvement in performance, then the inference is for super capacity processing (Figure 5c). If we think of a variation in workload as being accomplished by presenting increasingly more information about a face, then the notion of configurality would seem to be at odds with both limited and unlimited capacity processing. Instead, having more information (i.e., more of the holism) should lead to improvements in performance, suggesting that configurality should be associated with super capacity processing.

Figure 5: Graphical intuition of a system's behavior under different capacity bounds: limited capacity, unlimited capacity and super capacity. The total system's capacity resource remains the same across all conditions. (a) In the limited capacity case the total capacity is divided between two channels. (b) In the case of unlimited capacity each channel uses the total capacity. (c) In the super capacity case, the capacity devoted to each channel exceeds the total system capacity. The additional capacity could stem from some third part agent in the system. Note that an increase in channel capacity produces faster processing for that channel.



Methodology and Tests

Several tests have been designed in order to test the four fundamental properties of mental processing (J. T. Townsend & Ashby, 1983; Townsend & Nozawa, 1995). **The systems factorial test (SFT)** helps us uncover the architectural properties, distinguishing between serial and parallel processing, as well as determining stopping rule. It can also indicate a violation of the assumption of independent feature processing. The statistics that were derived to test processing properties are the *mean interaction contrast* and the s*urvivor interaction contrast* (Townsend & Nozawa, 1995). The latter is diagnostic function that is obtained after the analysis of survivor distribution response time functions.

Also another experimental manipulation of workload should provide data in order to test the system's capacity. The appropriate test would be inspection of the capacity coefficient, a **ratio of integrated hazard functions** for both comparison conditions (part- and whole-face). This test is based on a comparison of the total amount of work performed on separated face parts with respect to the total amount of work performed on a whole face.

A. Systems factorial test (SFT): Critical tests for architectural properties (additive factor method and systems factorial technology)

The additive factor method was designed in order to test the presence of serial short-term memory processing, when processes are independent and selectively influenced by their respective experimental factors (Sternberg, 1969). In order for the system to exhibit additivity, experimental data should show linearity, with the absence of interaction between manipulated factors. Sternberg (1969) proposed that mental processes are serially organized, indicated by the linear function between reaction time and number of elements in the memory set (search set). If the load is increased by one element, a constant amount of latency is added to the total reaction time (around 40ms). This time was recognized as the time needed to mentally scan a single item.

In order to test this model, Sternberg used an ANOVA design, combining sets of 2-3 experimental factors in a series of experiments (Sternberg, 1969). Findings indicated absence of interaction between experimental factors, thereby suggesting that factors operated in an independent manner. The independence between factors is realized through selective influence.

Note that theoretically it is possible for processes to affect each other not only directly over external factors (selective influence) but indirectly through mutual connection between processes (stochastic dependence).

Sternberg proposed that possible interaction between factors is due to a failure of the selective influence assumption rather than change in architecture (e.g. a change from serial to parallel processing). If selective influence for some experimental factor does not hold, then that factor could affect two or more cognitive processing stages.

A statistic that has been used to describe possible interaction between two factors with two levels could be presented as follows:

$$M_{IC} = RT_{ll} - RT_{lh} - (RT_{hl} - RT_{hh}) = RT_{ll} - RT_{lh} - RT_{hl} + RT_{hh} \qquad \text{equation 1}$$

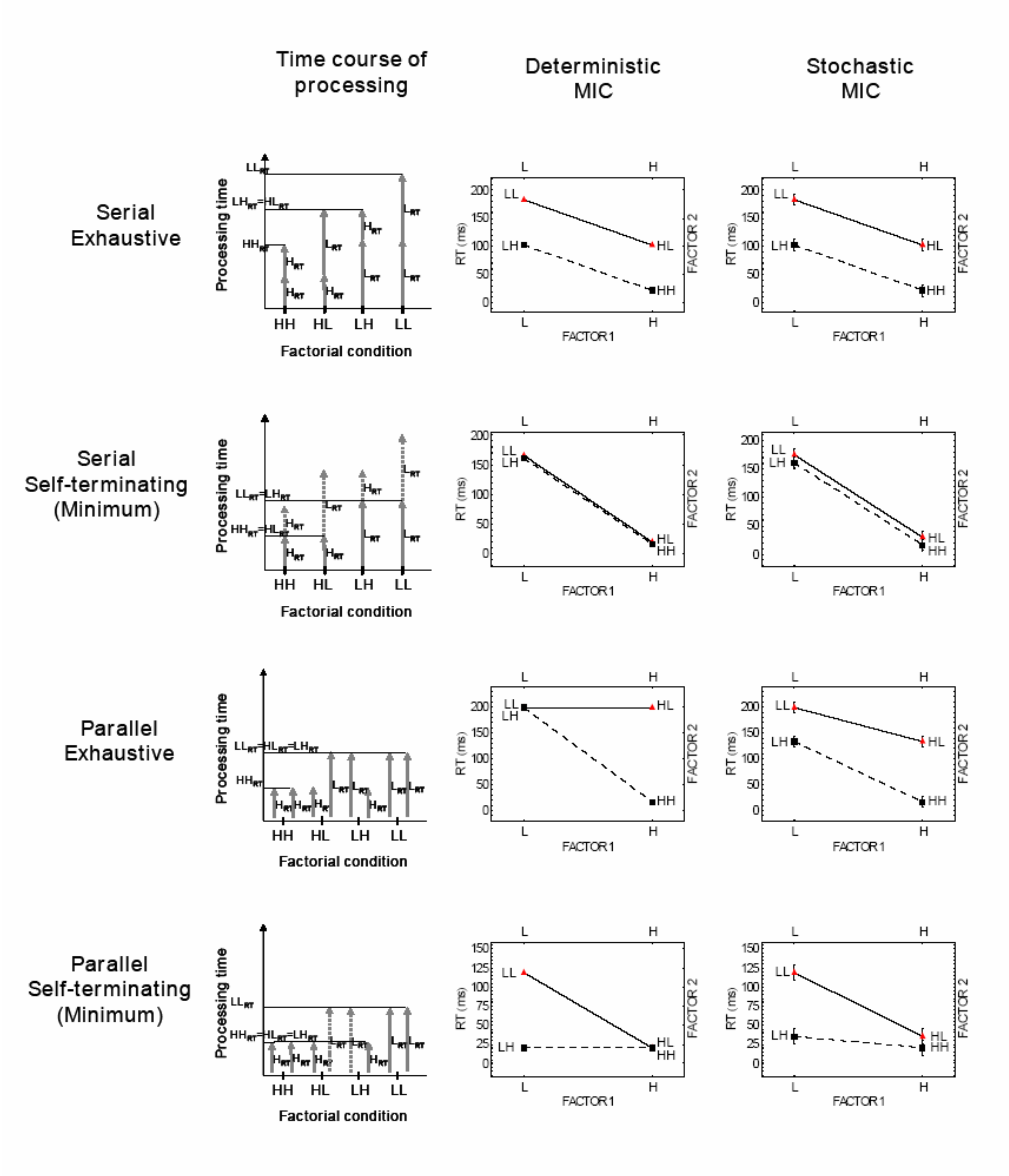
This is known as the <u>mean interaction contrast</u>. This statistic is obtained by the double difference of mean RTs associated with each level of separate experimental factors. In this case, 2 x 2 factorial conditions. The subscript denotes the values of the salience levels of each factor (h=high, l=low), while position of each subscript value indicates first or second factor. Here low indicated some manipulation that yields slower processing, high for faster processing, for example target brightness level. So, RT_{lh} indicates mean response time for experimental condition when first factor was on low salience/processing rate, and second factor was on high level. If additivity is observed in a data set then M_{IC} is identical to zero.

The limitation of the additive factor method was is inability to deal with architectures that are not serial - such as parallel or other systems. Schweickert offered a deterministic model based on graph theory, which showed how different systems could behave under assumption of selective influence (Schweickert, 1978). Most of the restrictions of the Schweickert theory were lifted by his subsequent work with Townsend (Schweickert & Townsend, 1989; Townsend & Schweickert, 1989). *The systems factorial technology* has been developed in order to test properties of mental processes organization under different architectures (Townsend, 1983; Townsend & Nozawa, 1995). The authors showed that possible interaction between factors, or in other words mean interaction contrast that is different from zero (negative or positive), could be explained by change in architecture properties and stopping rule (Figure 6) rather than failure of selective influence. So, $M_{IC}<0$, or underadditivity is typical prediction of a parallel exhaustive processing system. When $M_{IC}>0$, or overadditivity is observed, it is associated with a parallel self-terminating processing (see Figure 6). In contrast to

Sternberg's proposal (Sternberg, 1969), departure from additivity suggests that a system probably conducts parallel search that could be either terminated on finding a target or was performed for all elements. However these proposals were restricted to systems in which processes are stochastically independent and were selective influence holds.

In addition to the M_{IC} statistic, a more refined test has been developed based on survivor functions (Townsend & Nozawa, 1995). An important novelty was that a new assumption has been introduced – that selective influence operates on the distributional level. Note that a corresponding assumption of additive factor methods was that selective influence operates on mean level, that is - ordering of the means. But it is known that mean ordering allows the lowest level of statistical inference (Townsend, 1990). Much higher power of inference on behavior of one system is obtained when experimental manipulations are observed on distributional level, as ordering of CDF or Survivor functions implies an ordering of means, but not *vice versa*.

Figure 6: The time-course of processing of two items (left column), the corresponding deterministic (middle column) and stochastic (right column) mean interaction contrast (MIC), across different architectures and stopping rules (rows). The time course of processing depicts the change in total processing time for different factorial conditions (HH, HL, LH, LL) for different architectures. Each upright bold arrow in the graph corresponds to total processing of one unit (in the left column), that could be at the H (high) or L (low) level. A dotted upright arrow indicates a process that possibly did not complete because the processing terminated on a completion of a previous process. The deterministic MIC, in the middle column, represents the duration or the sum of process times (as indicated on the y-axis in the first column). Note that we are not able to directly observe the deterministic MIC in experiments because in real system processing, components will add some variability or noise. The stochastic MIC is an observable measure and is obtained when some variability or noise is added to the overall processing. Error bars around each mean condition represent the standard error statistic (added here arbitrarily for the sake of presentation).



The survivor function is a probability function that gives the probability that one component has survived some time. In our case, it measures processing that is not finished at time ‘t’. It is very easy to obtain survivor function from response times, as S(t)= 1-CDF(t) (cumulative distribution function), and most statistical programs nowadays have an option for calculating either the CDF or survivor function from a data set.

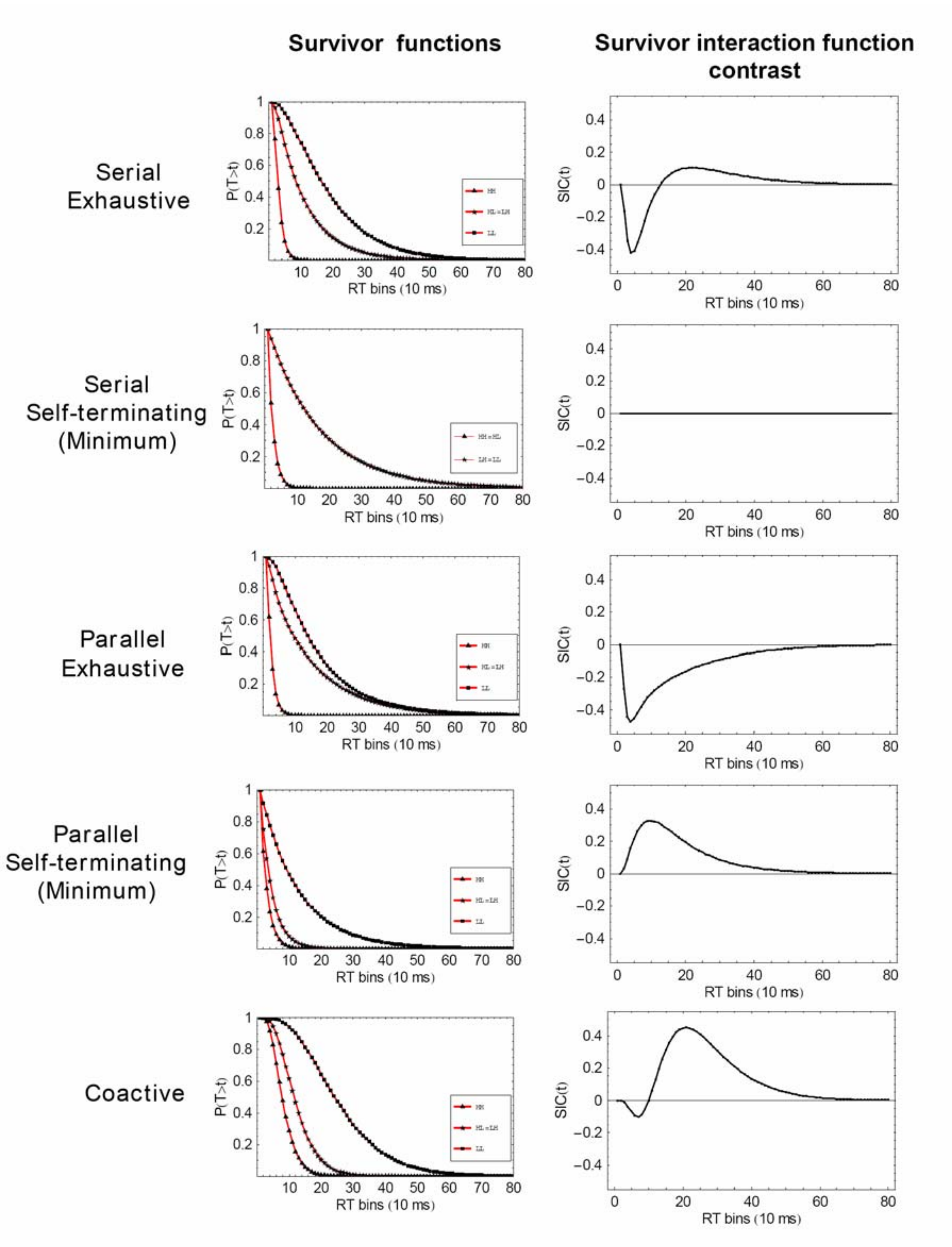
The *survivor interaction contrast function* is defined as (Townsend & Nozawa, 1995):

$$S_{IC}(t) = S_{ll}(t) - S_{lh}(t) - (S_{hl}(t) - S_{hh}(t)) \qquad (2)$$

Subscripts are defined in the same way as for MIC. There is specific signature of each stochastic processing architecture and stopping rule with respect to the shape of the $S_{IC}(t)$ function. For example, the parallel exhaustive model function is negative for all time, while serial exhaustive processing S_{IC} function is first negative and then becomes positive (ans S-shaped function). In Figure 7 we presented SIC functions for parallel, serial and coactive architectures combined with different stopping rule.

There is a straightforward mathematical relationship between M_{IC} and $S_{IC}(t)$. When $S_{IC}(t)$ is integrated from zero to infinity, it will return the exact value of the M_{IC}. This is well known mathematical property that integration of survivor function returns the expected time for that variable, which is the mean value. When $S_{IC}(t)$ function is integrated each term in Equation 2 is separately integrated because of the linearity of integration. Therefore it gives exact value of M_{IC}. Thus, the exact value of the M_{IC} is equal to the area of the $S_{IC}(t)$ function.

Figure 7: An ordering of joint survivor functions for different factorial conditions (HH, HL, LH, LL) (left column) and the survivor interaction contrast (SIC) (right column) across different architectures and stopping rules (rows). Note that each SIC function is calculated using $SIC(t)=S_{ll}(t) - S_{lh}(t) - S_{hl}(t) + S_{hh}(t)$. Each joint survivor function on the right hand-side is estimated from data (displayed in the left column). Note that each combination of architecture and stopping rule exhibits a unique SIC function. The shapes of these different SIC functions are independent of the form of the probability density function.



While the M_{IC} is single number, the $S_{IC}(t)$ is a function of time. It is important to note that different shapes of the $S_{IC}(t)$ could produce the same M_{IC} value. So it is obvious that the survivor interaction contrast is more rigorous test because it demands the exact shape to be matched for a particular architecture with stopping rule defined.

Information may be initially processed in parallel, and the outputs of this processing may be pooled or combined into a single "channel" of information. In this case, processing would be characterized as a special form of a parallel processing architecture known as coactive processing (e.g., Diederich, 1995; Miller, 1982, 1991; Mordkoff & Egeth, 1993; Mordkoff & Yantis, 1991, 1993; Townsend & Nozawa, 1995). Note that the coactive architecture assumes that all units were processed (exhaustive rule).

B. Capacity

This concept was investigated early in information processing research . Recently, it has received even more attention. The notion of capacity refers to a system's response to change in workload. If the cognitive system slows down when we increase the workload, then it is considered to be limited capacity. If it is not affected it is unlimited capacity. Of special interest to us, is the idea of super capacity, where the system speeds up processing with increasing informational load (Figure 5). This hypothesized behavior could be used to explain some facilitation effects that occur during perception of objects with good form, or gestalt processing. For example, some investigations reported a face superiority effect in recognition where it was demonstrated that a face feature is more accurately recognized in the background of the study face then when a foil face was used.

One important thing to observe is that capacity can be considered at the individual item level (single process) and also at the level of global processing (whole system). Note that in can occur that while processing on a single item is of unlimited capacity, which means that this item will be processed at the same speed regardless of the total amount of work to be done, the whole processing system is not necessarily unlimited. For example, a parallel model with independently processing elements, and an exhaustive stopping rule, will show an increasing mean reaction time as a function of the number of processed elements, under unlimited capacity at the single element level. By changing the stopping rule to self-terminating, it will exhibit overall unlimited capacity, that is, it shows a flat function of reaction time as a function of the workload. And further, if the model uses minimum time processing, then it will show decreasing mean reaction times as a function of the workload.

So capacity is closely related to manipulations in architecture, stopping rule and/or interdependence between processing units. In order to measure capacity, we will use capacity the coefficient (Townsend & Nozawa, 1995). The capacity coefficient is a ratio between cumulative hazard functions for holistic case (when all elements are processed together), and sum of cumulative hazard functions for separate elements of the whole unit.

A hazard function is conditional probability function which is defined as a ratio between density and survivor function: $h(t)=f(t)/S(t)$. When integrated over time, it gives the integrated hazard function, $H(t)$, which is a coarser measure but is more stable than the hazard function $h(t)$. This statistics is considered measure of total amount of energy expended to complete the task by some time *t*. The hazard function $h(t)$ can be viewed as power, while the integrated hazard function as amount of work done. The integrated hazard function can be obtained directly from the data using a logarithmic transformation of the observed survivor function H(t)=-ln(S(t)) (Townsend & Ashby, 1983; Townsend & Nozawa, 1995; Wenger & Townsend, 2000).

The capacity coefficient is calculated as a ratio between work done with full workload with respect to the sum of total work done on a partitioned workload. We denote $H_{AB}(t)$ as the work done on two components at the same time, and $H_A(t)$ and $H_B(t)$ the work done on both components separately.

For OR processing case (self-terminating), the capacity coefficient for two elements is defined as

$$C_0(t) = \frac{H_{AB}(t)}{H_A(t) + H_B(t)}$$

For AND processing (exhaustive terminating rule), we will define a new function analogous to H(t) as a K(t)=ln(F(t)), where F(t) stands for the cumulative distribution function . Note that here the different stopping rule demands change in the calculation our function. The capacity coefficient for two elements is now defined as

$$C_a(t) = \frac{K_A(t) + K_B(t)}{K_{AB}(t)}$$

If both capacity coefficients (Co(t) and Ca(t)) have value >1, then processing is consider to be super capacity; if C(t) =1 then processing is unlimited capacity, and if C(t)<1 then processing is limited capacity.

The base model for both the Ca(t) and Co(t) is the unlimited capacity parallel model (UCIP) which is why C(t)=1, is unlimited capacity.

General hypotheses on holistic/configural processing

So now state our general research hypotheses concerning organization of mental processes in configural/holistic perception:

1. **Coactivation or parallel dependent processing**: If processing is holistic then double factorial experiment applied on face recognition will reveal parallel or some form of coactive processing (revealed by the SFT test). It is also tenable to assume that processing could exhibit parallel dependent signatures.
2. **Exhaustiveness**: We could expect exhaustive processing considering that holistic object would probably take advantage of all its parts in perception. A typical signature of exhaustiveness is exhibited differently in parallel and serial models (revealed by the survivor interaction function). Exhaustiveness is not mandatory, since all parts do not have to contribute of the object perception, especially in OR case.
3. **Super capacity**: workload manipulation (part-face, whole-face) should indicate average faster, more efficient processing for whole faces. Also, the capacity coefficient will indicate super capacity $C(t)>1$ for the whole face.
4. **Stochastic Interdependence**: In holistic/configural processing there should be some cross-talk between channels (or units) of processing. A stronger assumption is that there should be facilitatory interactions between processing units. Processing on one unit (say John's eyes) should help processing of other units (say John's lips). In short we could expect facilitation between units of

perception to be realized through positive stochastic dependencies between processing times of those units. In a more radical form with extremely high positive interactions between units processed in parallel, it is possible that the system degenerates into a form of coactivation.

General experimental design

This study was divided into two parts, using a between-subjects design: OR and AND experiments. For each part, participants must use a different stopping rule in order to correctly categorize a set of faces. In the OR experiment, recognition of a single face property could yield correct categorization, while in the AND experiment, participants must process two manipulated face properties to categorize all faces. A between-subject design was employed due to the very large size of each experiment, which for most participants, involved approximately 24 1-hour sessions.

Two configural face properties were manipulated: the eye-separation, and the lips-position in a face. The features itself were not altered. Thus, the main manipulation is based on altering the distance between the facial dimensions or features (for detailed methodological issues see Rakover, 2002). These are considered to be manipulations of configural face properties (e.g. Diamond & Carey, 1986; Leder & Bruce, 2000; Tanaka & Sengco, 1997)

Each study had two parts: **the learning phase** and **the test phase**. The goal of the **learning phase** was to train participants to correctly recognize faces separated to in two groups. The progress of each participant was monitored by inspection of both mean reaction times and errors as a function of learning sessions. In order to reveal the properties of architecture, stopping rule, capacity and dependency, results were tested by the SIC, MIC, and capacity tests, for each session.

The goal of the **test phase** was to observe possible changes in any of the processing properties (architecture, dependency and capacity) when a face configuration is changed. The goal of this phase is to seriously impact previously learned configuration of features by disrupting it in several ways:

(a) By removing all parts of the face that were not manipulated and observing only two dimensions in isolation (the featural test condition).

(b) By embedding both learned dimensional properties in a new facial context (the configural test condition).

Performance on each of the two configural disruptions will be fastest and yield fewer errors in the standard test condition, that contains regular old faces.

Our research hypothesis is that categorization would be the fastest, with fewer errors, in the standard test condition, than in both the (a) and (b) conditions.

Compared to similar studies of face perception, (e.g. Tanaka & Sengco, 1997), this study should provide a more rigorous test of processing architecture. Of special interest are the processing characteristics in the second testing phase, because we expect processing to change from holistic to one that is more feasible under analytic processing. With configural disruptions we expect that architecture could change to slow serial processing, or parallel limited capacity, and that the system will exhibit independent feature processing.

The confounding effect of novelty

Important note about the learning phase: In the test phase, introduction of a novelty (part of face or background of a face) to an old face should be performed by using features from a face that had been previously learned. In short, in the first part of the experiment (the learning phase) a participant will learn two distinct groups of faces. In the test phase configuration of faces from the first group will be disrupted by using features from the second face group.

A rationale for this is that the use of completely new faces, or its parts, might confound the disruption of the previously learned face. Changes in performance could be based on introduction of a novel stimulus rather then just manipulations that disrupt configuration only. Note that if the novel part, say a new nose, is put in to a previously learned face, this could produce two effects: novelty effect and configural disruption effect.

If completely new face part is used for a disruption, there is some theoretical chance that the system will react to the novel stimulus, probably because attention could be attracted to the new and previously unseen (not learned) part. We hypothesize that independent analytic processing could be more affected by the novelty than by the configuration disruption, in contrast to holistic processing. In fact, analytic processing could slow down due to the additional processing of a novel part. So, the potential source of a confound lies in possibly different effects of a novel disruptive feature and a previously learned disruptive feature, on both analytic and holistic processing.

In order to avoid this potential confound, in a separate learning block, participants will learn a second face, in contrast to the categorization for the first face (in the AND, OR tasks). The task for the second face will be to learn to associate a different name with 4 different faces.

Blocked and mixed face-context conditions

In the *blocked face-context condition* all face-stimuli in a block of trials were either whole faces, or part-faces mixed with whole faces. The reason for blocking whole-faces is our concern that mixing part-faces, or transformed faces, with whole-faces could increase the usage of analytic processing on all stimuli. At this stage we don't reject the idea that a face representation could contain both whole-face and part-face representations, but we suggest that the whole-face could be considered as a super-ordinate entity in contrast to part-face representation. We could also imagine whole- and part-face based processing strategies as the opposite ends of one continuum that spans from part-based to a more holistic representation.

However we can not avoid mixing both whole- and part-faces. This will be done in one separate block needed to investigate the system's capacity. In the *mixed face-context condition* both the part- and whole-faces are used in the same experimental block.

The *experimental face-context effect* is tested when the blocked whole-faces or blocked part-faces are compared in mixed and blocked face-context conditions. For example, we compare reaction time needed to recognize whole face in the blocked condition to the mixed condition. A non-significant experimental face-context effect, i.e. suggests that the holistic and part-based representations do not share common cognitive structure.

Survivor interaction contrast function and Capacity function

In this part we try to provide a more detailed description of how all the tests (MIC, SIC and capacity) are calculated at different experimental stages, and we aim to describe their relationships, in terms of the conditions used to calculate them.

First, note that the single operative unit in this study is a face. Single whole-faces are used during learning sessions in order to facilitate holistic encoding. However not all faces are presented as a whole: the masked-faces appear like whole-faces but some face regions vary with respect to brightness. The masked faces appear like a partially hidden in the shade of a tree (see Figure 8).

Figure 8: An example of two masked faces.

Traditionally, in order to explore holistic encoding, a part-based face is used, with the parts or surrounding being removed. We define the masked faces as a complement to a part-based face-stimulus given that it does not convey full information from the face (see Appendix A). The masked faces are used in one portion of the learning section of the experiment in order to test the level of face-representation capacity. Another part-based face is the feature-face, which appears in the second part of the experiment (the test sessions). The participants observe only two configural face-properties of interest, which are the eye-separation and lips-position. But they are instructed that the isolated configural features, always presented together, belonged to one of the faces they learnt previously.

In order to apply the MIC and SIC tests, which are ubiquitous tests for the system architecture and stopping rule, we need four single faces, with two face properties factorially combined at two level of activation/salience. So both the MIC and SIC tests can be only applied on a part of a face-space used in the experimentation. They are not single-point-face sensitive tests. Further, in our experimental designs, in both the AND and OR conditions, only one set of faces could be used for the architecture test, given that they are the set formed by factorially combining the distance between the eyes and the lips.

In contrast, the capacity test can be applied on all faces in the experiment, provided that for each single whole- or feature-face the appropriate masked-face is used for comparison (see the experimental design section for more details). The capacity test could be applied on a set of four faces used for the SIC and MIC tests, by collapsing across a category of faces.

Definitions of capacity coefficient function (CCF)

As we defined above, the CCF is a ratio between an index of processing of a whole object and the sum of indices of processing for separate parts of that object. Although it has a different mathematical form for both the AND and OR paradigms, the logic of calculation of the CCF is practically the same in both cases. Putting it in more appropriate words, we could say that the CCF reflects properties of the system characterized by the ratio of a measure of total mental work done on a whole-face, with all the face features presented in a biologically appropriate arrangement, and the total amount of mental work for each complementary masked face.

Our experimental paradigm allows us to measure the CCF in several different ways, because both indices (numerator and denominator) can be obtained from different experimental conditions. Each new CCF has a particular, important role in understanding of the system's behavior. In order to describe the properties of the system related to a concept of the capacity we will define several ways to calculate the capacity coefficient function:

In terms of the experimental block context: we will provide a different definition of the CCF depending on whether faces and parts are mixed together, or presented alone. In the two experimental conditions: the *blocked face-context* and *mixed face-context.* Note that in our study both conditions are utilized. However, in the blocked face-context condition, only whole faces are presented, while in the mixed face-context condition both whole faces and part-based (or masked faces) are presented. The mixed condition was called the capacity test, and was designed for investigation of the capacity properties of processing of encoded images. However, two outcomes regarding the speed of recognition of wholes and parts are formulated. The first is based on face-context free effect suggesting that there should be no difference in reaction time for whole faces in both the mixed and blocked conditions. The second outcome suggests a dependency between the encoding of whole faces and its parts, which will be observable as (usually) slower processing of whole faces in the mixed condition relative to the blocked condition.

Therefore, we will calculate both CCFs, one that will use an index of processing for whole-faces the blocked condition, and one that will use whole-face index from the mixed condition. The index of part-based processing will be identical in both cases and it will stem from the capacity-test part (mixed blocked condition). We will denote the first measure *the whole-blocked CCF* and the second measure *the whole-mixed CCF*.

The logic behind this manipulation is to maximize the identification of holistic face-processing properties. Usually, we would proceed with the second CCF only. But we find evidence that blocked whole faces elicit faster responses than when they are mixed with part-based stimuli, then that could obscure our identification of the gestalt processing properties by neglecting this source of information.

The second way to define the CCF is to alter the relative points in the learning process. In terms of the learning process we will define the *absolute learning CCF* and the *relative learning CCF*. The absolute learning CCF reflect changes in the capacity coefficient by measuring the amount of work done on whole faces has changed relative to the first day/session of training. More formally, the capacity index for the whole face will change for different learning sessions, while the capacity index for the part-based encoding will be kept constant and taken from the very first session. The absolute learning CCF will about the total amount of change of the system's capacity from the very first session, for the encoding of whole faces only.

The relative learning CCF reflects the possibility that learning could also change the processing of the face parts over each experimental session. More precisely both indices of the whole-face encoding and the part-based encoding will change as a function of the experimental (learning) sessions. This CCF provides information regarding the relative amount of learning specifically for that session, by comparing the amount of work done on whole faces and part-based faces, for that session.

Both indices could show similar trends, but it is possible that they indicate different capacity patterns over the course of learning. In particular, it is possible that the absolute learning CCF shows an increasing trend of the capacity index over the session that could correspond to more holistic encoding of the whole face, while the relative learning CCF shows no change due to part learning. Alternatively, it is also possible that over the course of learning, participants start to pickup more information from parts and treat them as holistic object themselves, which will be indicated by a reduction in value of the relative learning CCF as a function of learning sessions. So, given that both capacity indices could provide different information about the processing of whole and part faces, it is useful to utilize the both.

Definition of learning

Although the term learning is defined differently depending on the area and the focus of research, we will tie this term to changes in the capacity. In this study the capacity is revealed by the capacity coefficient index that is operationally defined through the amount of work done. In fact, it is defined as a ratio of work done when a whole face is processed relative to the work done when part of a face is processed. The learning aspect refers to the change of both indices, whole and part-based work, across sessions.

However, the learning process itself could have additional consequences or impacts on the face recognition system. It could enhance the status of a memory representation (e.g. Sergent, 1984) and/or change other properties of the cognitive system, such as the architecture, stopping rule or interdependency (e.g. Wenger & Townsend, 2000, 2001). Further learning could occur through automatization (e.g. Czerwinski, Lightfoot, & Shiffrin, 1992; Logan & Stadler, 1991; Shiffrin & Schneider, 1977).

The capacity coefficient function can not easily or directly prove, reveal or even disconfirm some aspects of learning. Rather, it can indicate whether learning occurs and to what extent it alters processing efficiency, and very importantly, it provides a better statistical description of the learning process in contrast to utilization of RT means as a measure. In order to determine which property of the face encoding system has changed, more information is needed. Therefore we will pair the capacity coefficient together with the test for the system's architecture and stopping rule, revealed by the MIC and SIC tests.

With this combination of tests, the learning process can be characterized in more detail. We will discuss the expectations of the system's behavior in later sections that describe specific research hypotheses. For now, we will only suggest that the learning could occur through the development of a better memory representation (face representation), through a decrease in noise in the representation, or through enhancement of memory trace strength, and/or through a change of the system's architecture when face comparison occurs (say, from slow serial to fast parallel).

<u>Further definitions: 4 capacity functions</u>

So, from the methodological point of view we could investigate properties of the capacity along two important dimensions: mixing or blocking of whole faces with its parts (mixed-blocked), and amount of learning with respect to the beginning of a study (absolute learning) or amount of learning for each session. Combination of these dimensions allows us to calculate four different capacity coefficients:

1. *Absolute learning whole-blocked CCF*: indicates the total amount of learning by comparing the amount of perceptual work done on a whole face during the most recent session to the amount of work done on a part-face calculated from the

initial session. We expect this CCF function to show an increasing trend during learning, mainly due to more efficient learning of whole-faces better learning the faces over session trials. Given relatively low discriminability between the faces to be learned we expect that participants will show low or limited capacity index at the beginning and that the CCF will tend to increase toward super capacity over the course of learning. Note that this function is not affected by the possible learning of face-parts, because part-based measure of the capacity ratio function (for both AND and OR conditions) is taken from the initial learning session and is kept constant, while the whole-face measure changes as a function of learning session.

2. *Absolute learning whole-mixed CCF*: this function is identical to the previous one (1), except that the whole-face measure stems from the part of experiment (Figure 9, the 2nd part) where both whole- and part-faces were mixed together in one block. This block was run at the end of each experimental session in order to test the capacity properties of the current learning session. We expect that this index will increase as a function of learning sessions. It is possible that this measure will be of lower magnitude when to (1), because mixing together whole- and part-faces may produce slower processing of whole-faces than when they are blocked. This contextual hypothesis stems from our pilot experiments.

3. *Relative learning whole-blocked CCF*: this function is calculated such that both the whole- and part-faces responses are taken from current learning session, while the whole-face part stems from the blocked face condition, and the part-face observations are from mixed-block condition. We expect that this function will

not necessarily increase across the learning sessions, but could show an initial increase and then decrease later, depending on amount of learning for part-based faces. In short, a non-monotonic trend is expected. The reason for this expectation is the possibility that participants could learn part-faces as well as whole faces, using a similar gestalt-learning mechanism. The increase in efficiency of part-based information recognition could equate to the learning of the whole-faces.

4. *Relative learning whole-mixed CCF*: this function is identical to the previous function (3), except that the whole-face measure stems from the experimental condition (Figure 9, the 2nd part) where both whole- and part-faces were mixed together in one block. We expect that the functions (3) and (4) (over different sessions) are also of lower magnitude when compared to either (1) or (2).

Method section intro

The goal of the **learning phase** was to monitor any changes in the architecture and capacity during category learning. Thus, each experimental session consisted of two parts: in the first part, the test for architecture was applied, and in the second one the capacity test was used. Additionally, the test for architecture was incorporated into the second portion. The only constraint preventing us from using the architecture test in the second part was the relatively small number of observations per condition collected; more trials are needed for an accurate test. There is also a possible unwanted consequence of architectural change in the recognition of whole faces when both whole faces and part faces are mixed together. In fact, our analysis did show that whole faces are recognized slower when blocked together with part-based (masked) faces.

During the learning phase, participants learned to recognize two sets of faces. They had to press mouse buttons with different index fingers for each category response.

On the first session participants were informed which dimension are important: eye-separation (wider or closer) and lip-position (up or down). The reason for providing participants with important information concerning experimental manipulations stems from the fact that we want participants to reach errorless performance in both tasks. Namely, our pilot investigations showed convincingly that without initial help, participants could not reveal the important face dimensions within the first couple of sessions in the AND task. Providing a precise description of face-space helped participants to establish more accurate and reliable performance early in the training. In order to equate manipulations for both the OR and AND cases we had to provide the same help for the OR case although it was not as necessary as in the AND condition.

OR CONDITION

Experiment learning phase

Method

Participants

Six participants, 3 females and 3 males were paid for their participation.

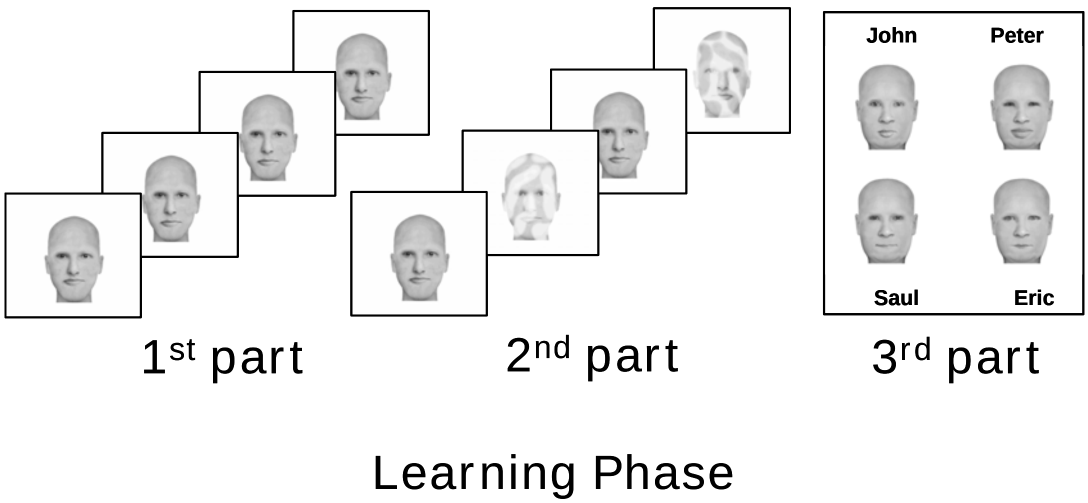
Materials

In the OR categorization task, participants had to decide whether a displayed face belonged to the group of gang members or was the hero face. In the first part of each session, participants observed only whole-faces. The goal of this part was to investigate organization of mental processing by application of the SFT tests (the MIC and SIC). In the second part participants observed mixed whole-faces and masked-faces. The goal of the second part was to test the capacity of the system (the capacity test). In the third part participants switched to a complete identification task, and they had to learn to associate each of 4 names with a particular face. The goal of this part was for participants become familiar with face features that will be used in the test phase of experiment. Thus, participants completed all three parts in each learning session (see Figure 9)

In **the first part** we manipulated the following factors: (a) face category, (b) feature configuration, and (c) feature saliency.

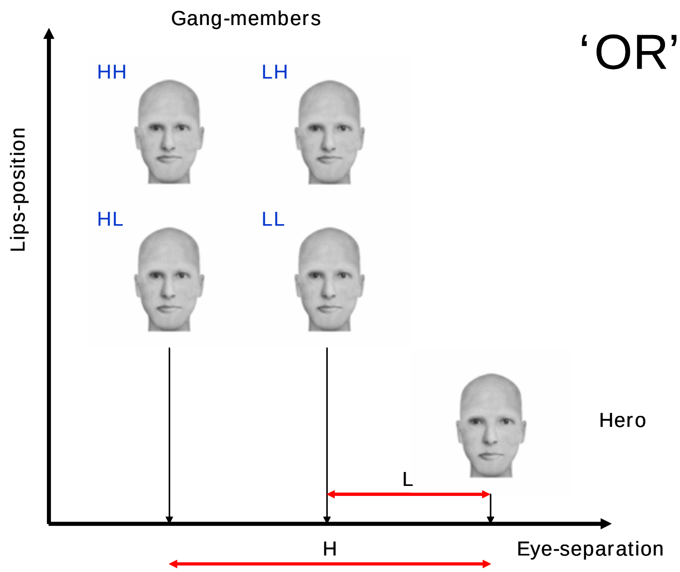
For (a), faces could be either the hero a gang member. Four gang member faces were designed based on manipulation of (b) which involved changes to either the distance between the eyes or the height of the mouth relative to the nose. Finally, the saliency or detectability of the configural changes was factorially combined with (b); the resulting four faces for the set of gang members.

Figure 9: The learning session consisting of three different experimental parts. In the first part, we used only whole faces. In the second part, we combined whole faces from the first part with masked faces. In the third part, participants identified 4 individual faces. This structure of one learning session was common for both the OR and AND designs.



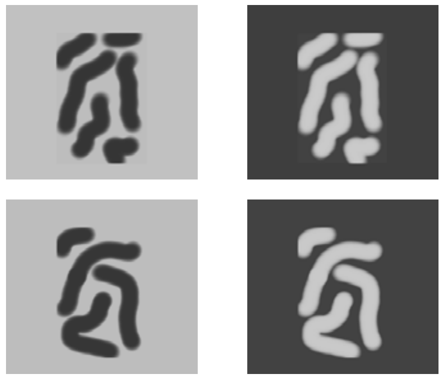
The saliency of features is defined by the marginal proximity of each face-feature projection onto two-dimensional face space (see Figure 10) with respect to the other group members. In this case the projection, or marginal value, of each face-feature for each gang-member is either close or distant with respect to the projection of same feature from the hero. Factorial combination of the face-features (b) and the salience level (c) produces four combinations: HH, HL, LH and LL. First letter denotes the saliency of the eye-separation, while the second letter denotes the saliency of the lips-position. For example, the condition HH defines one gang-member who is the most distant from the hero face, in the designed face-space, hence the difference in eye-separation and lips-position are easiest to detect. Here notation H (high) produces faster recognition than the L (low) condition. The saliency is inversely related to the similarity between faces, so a low salience feature on a gang-member face is more similar to the hero face feature than a high salience feature.

Figure 10: Two dimensional face-space defined by the eye-separation and lips-position. The saliency of features is defined by the marginal proximity of each face-feature projection onto the axes with respect to the other group members. In this case the projection, or marginal value, of each face feature for each gang-member is either close or distant with respect to the projection of the same feature from the hero. Factorial combination of the face-feature and the salience level produces four combinations: HH, HL, LH and LL. The first letter denotes the saliency level of the eye-separation, while the second letter denotes the saliency level of the lips-position.



In the **second part**, in addition to the experimental manipulations (a), (b) and (c) from the first part another factor was used: (d) whole-masked faces. On each trial either a whole- or masked-face is presented on the screen in a random fashion (Figure 9). Masked faces are designed using two types of "worm-like" masks (Figure 11).

Figure 11: Two masks used in order to generate the masked faces are presented in the first column. In the second column, we showed brightness-inverted masks used to generate complementary masked faces.



Each mask was designed in Photoshop 7, and for each mask, a complementary brightness mask was designed (Figures 12 and 13), producing a total of 4 masks. Multiple masks were designed in order to increase masking variability and to prevent possible learning of particular features of a single mask that could eventually produce an unwanted search strategy. Masked-faces are generated by superimposing the mask over a face using the Mathematica 5 environment by calculating resulting the image by multiplying a face and mask picture. The inverse mask was used to obtain get an inverse masked picture of the whole face. Similar manipulations on other types of stimuli were used in the investigation of configural processing in fingerprint experts (Busey & Vanderkolk, 2005).

Figure 12: The design of the masked faces. Each face was multiplicatively combined with the original and brightness-inverted masks in order to generate two complementary faces. Simple summation of the faces in the top row will return the original face.

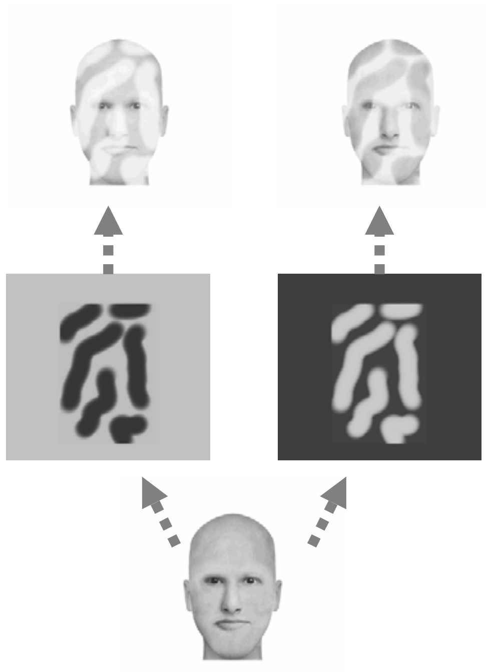
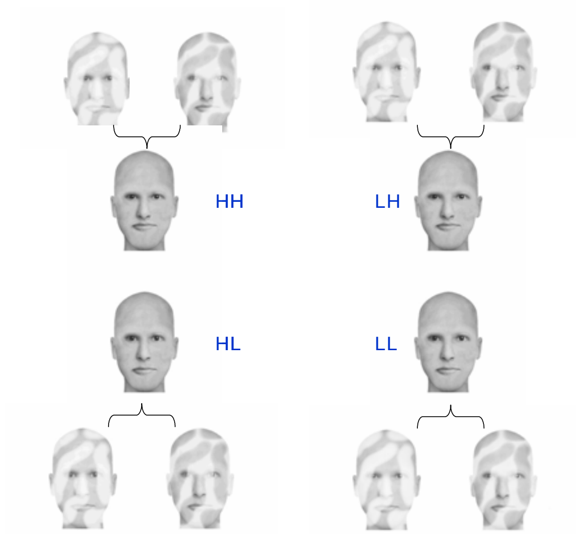
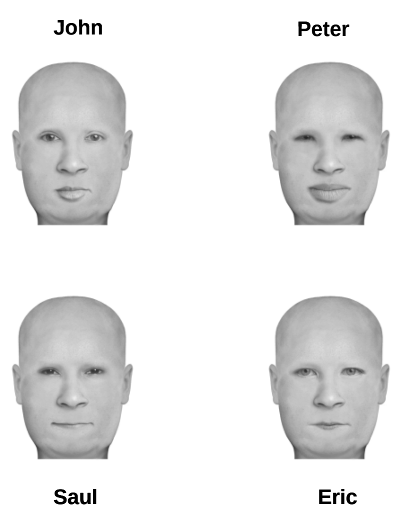


Figure 13: Example of one set of masked faces for the gang-member faces (in the OR condition), or Sharks faces (as they were named in the AND condition).



The third part of the experiment a complete identification task where participants had to learn to identify 4 faces (Figure 14). Participants were instructed to learn a second gang of four faces, where each individual was given a name (John, Peter, Saul and Eric). One face was presented on a monitor at time, and the task was to learn to recognize them. Participants responded by pressing one of four buttons on a numerical pad (1 to 4). All faces contained different face-features (eyes and lips) but they had same face background. The goal of this part was to implicitly introduce participants the face-background that would be used in the test phase of the experiment (next section). Again, the background of the face (everything but the eyes and lips) were common for all four faces and were not critical for differentiating between faces.

Figure 14: The four faces used in the complete identification task in part three of the learning session. Participants responded by pressing one of four buttons on a numerical pad (1 to 4). Press 1 if you see John, 2 if you see Peter, 3 if you see Saul, and 4 if you see Eric.



Design and procedure

Gang-members were presented on half of the trials, and the other half were the hero-face. A participant had to decide, by pressing one of the mouse keys with left and right index fingers, whether a hero or a gang-member was presented. RT was recorded from the onset of a stimulus display, up to the time of response. Each trial consisted of a central fixation point (crosshair) for 1070ms followed by a high-pitch warning tone which lasted for 700msec. Then a face was presented for 190ms. Upon incorrect categorization participants received an error message.

In the first part of the experiment, participants a received total of 600 trials. The hero-face was presented 300 times; gang members were also presented for 300 trials, with each gang-member face presented on 75 trials. In the second part, participants observed another 240 trials, of which 80 belonged to both whole-faces (divided equally by the hero and gang members), and the remaining 160 trials belonged to the two same groups but with different types of masks (80 trials for each mask applied on both groups). In the third part each face was presented on 75 trials.

A trial presentation order was pseudo-randomized within each session.

In each session, participants ran 5 blocks of approximately 200 trials each. All participants accomplished a total of 12 sessions during the learning phase of the experiment.

The participants were instructed to achieve high accuracy and to respond as fast as possible. For the analyses we aggregated two subsequent learning session sessions into one, therefore obtaining a larger number of observations per condition. So for the SFT tests, each factorial condition (HH, HL, LH and LL) possessed approximately 150 trials.

The reason for the data aggregation was primarily to get more observations in the capacity test part. For the capacity analysis, we analyzed approximately 160 trials for each integrated hazard function.

Results

Basic Mean RT Analyses

The GLM univariate analysis was conducted on both the gang-members faces and the hero face separately. However, we will focus our attention mainly on the analysis of the gang-member faces, given that they are factorially manipulated and allow for application of the SFT tests (the MIC and SIC).

The results from GLM analyses for gang-members, for each participant, are presented in Table 1.

Gang Members

First, we analyzed overall learning trends for the gang-member faces, for all participants: reaction time as a function of the effect of learning sessions (total number of sessions/2). We found that learning trends are all highly significant, (see Table 1) under the factor learning session). The decreasing trends of RT as a function of learning session are presented in Figure 15, combined with error rates across learning sessions (proportion of errors). Generally, it can be observed in Figure 15 that different participants exhibited either monotonic decreasing trend of mean reaction time as a function of the session that approaches some asymptotic level (Participants 3, 4, 5 and 6), or they exhibited inverse U-shaped trends (participants 1 and 2). At the same time, mean error levels exhibited a consistent decreasing trend as a function of the session (bars in the Figures 15). We can conclude therefore that Participants 1 and 2 exhibited a speed-accuracy trade-off, with fast reaction times on some sessions that resulted in a higher proportion of errors.

Almost identical trends across participants were observed in the hero condition (Figure 16).

Figure 15: OR condition, learning of the gang faces. Mean RT as a function of learning session, combined together with the error rate (proportion of errors), for all participants.

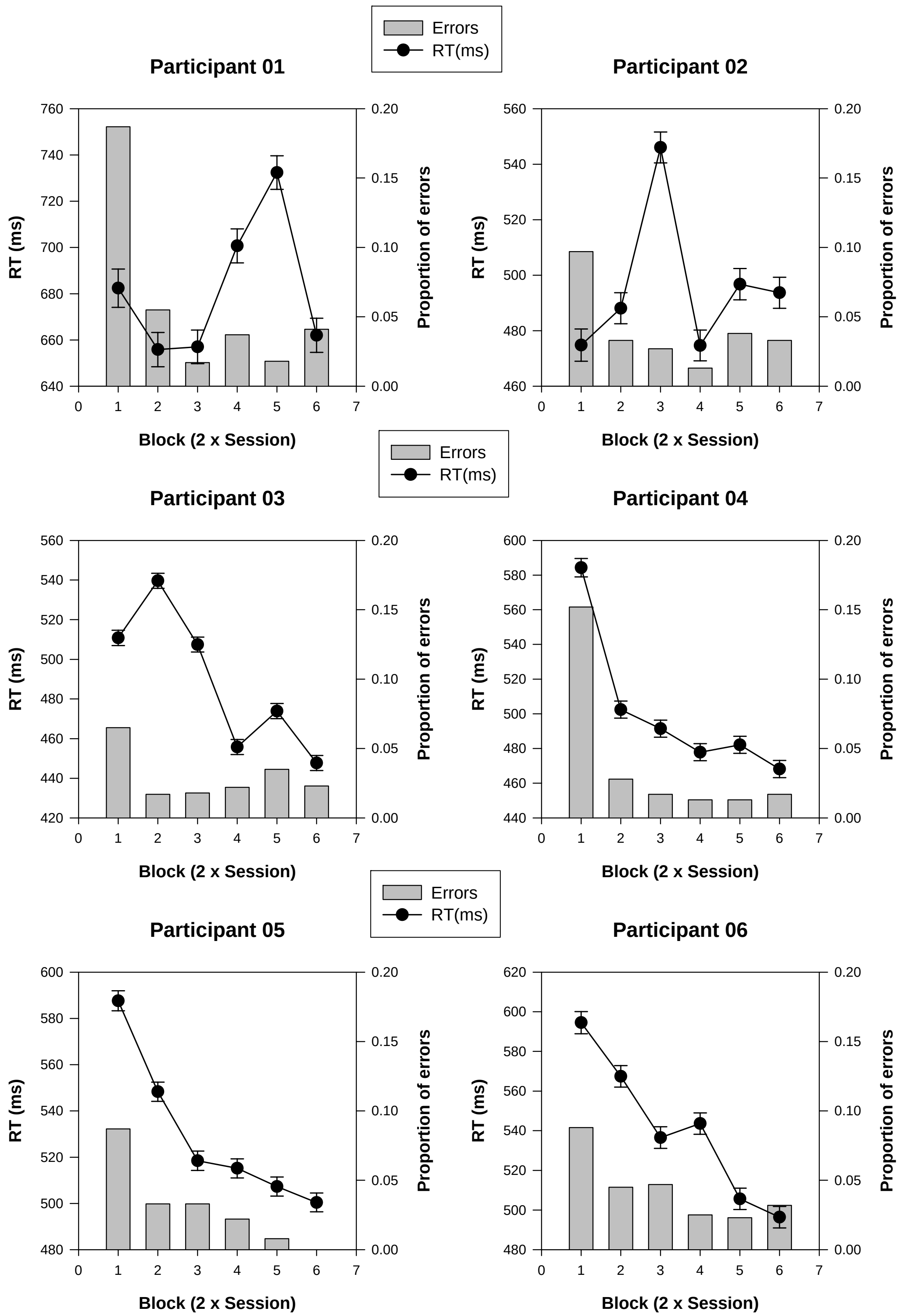


Figure 16: OR condition, learning of the hero face. Mean RT as a function of learning session, combined together with the error rate (proportion of errors), for all participants. Error bars on mean RT indicate standard error.

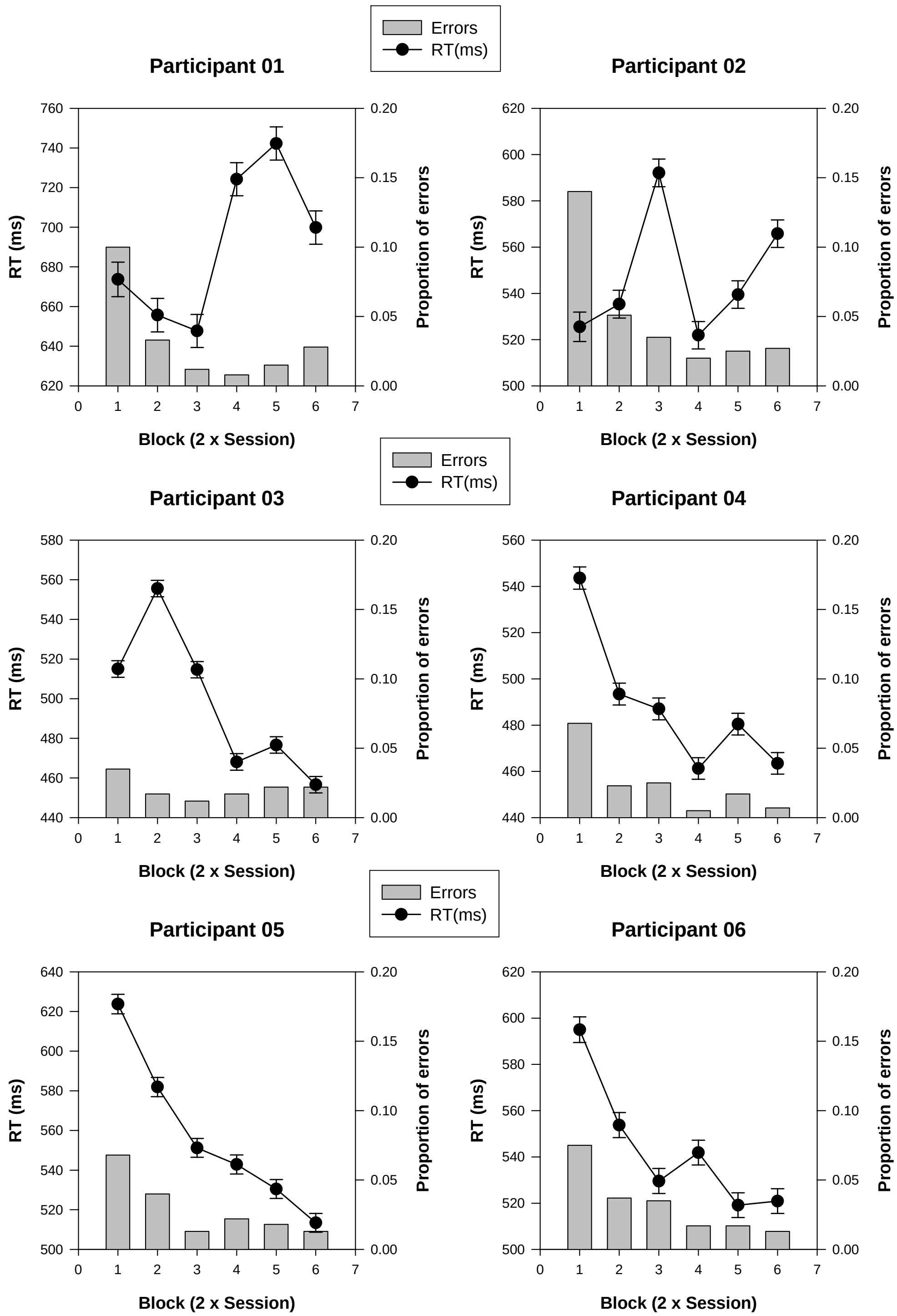


Table 1: The GLM univariate analysis was conducted on the gang-members, for different participants (SUB). The main effects and interaction terms are listed in the second column (*Factor*). Degrees of freedom are in the third column (*df*). The *error* row defines the degrees of freedom for the F-test error term, for that participant. Each F-test value (*F*) has two degrees of freedom: one from its corresponding row, and the other from the error row. The significance level is presented in the column *Sig.*, and the observed power for that effect is in the last column.

	Factor	df	F	Sig.	Observed Power		Factor	df	F	Sig.	Observed Power
SUB 01	Trial order	1	6.387	.012	.715	**SUB 04**	Trial order	1	11.846	.001	.931
	Eyes	1	496.589	.000	1.000		Eyes	1	377.966	.000	1.000
	Lips	1	564.643	.000	1.000		Lips	1	64.698	.000	1.000
	Learning Session	5	17.688	.000	1.000		Learning Session	5	63.203	.000	1.000
	Eyes x Lips	1	245.695	.000	1.000		Eyes x Lips	1	31.077	.000	1.000
	Eyes x Lips x Learning Session	5	3.523	.004	.920		Eyes x Lips x Learning Session	5	2.271	.045	.741
	Error	3362					Error	3430			
SUB 02	Trial order	1	4.089	.043	.525	**SUB 05**	Trial order	1	1.169	.280	.191
	Eyes	1	196.695	.000	1.000		Eyes	1	584.505	.000	1.000
	Lips	1	215.888	.000	1.000		Lips	1	307.153	.000	1.000
	Learning Session	5	21.247	.000	1.000		Learning Session	5	59.618	.000	1.000
	Eyes x Lips	1	87.142	.000	1.000		Eyes x Lips	1	154.230	.000	1.000
	Eyes x Lips x Learning Session	5	1.550	.171	.548		Eyes x Lips x Learning Session	5	2.004	.075	.678
	Error	3430					Error	3465			
SUB 03	Trial order	1	61.544	.000	1.000	**SUB 06**	Trial order	1	14.011	.000	.963
	Eyes	1	346.653	.000	1.000		Eyes	1	497.697	.000	1.000
	Lips	1	287.209	.000	1.000		Lips	1	265.203	.000	1.000
	Learning Session	5	90.033	.000	1.000		Learning Session	5	46.465	.000	1.000
	Eyes x Lips	1	121.497	.000	1.000		Eyes x Lips	1	142.552	.000	1.000
	Eyes x Lips x Learning Session	5	1.374	.231	.491		Eyes x Lips x Learning Session	5	5.593	.000	.993
	Error	3467					Error	3419			

As can be seen, both manipulated face feature properties, the eye-separation and lip-position factorially combined with feature saliency, exhibited significant main effects, for all participants, at the $p<0.01$ level. Thus, manipulation of configural face-feature properties produced significant perceptual effects. Both properties also exhibited significant change in detection over learning blocks (Eyes x Session and Lips x Session) for most of participants (Table 1).

Of the utmost importance, the MIC test, which was tested by the significance of the Eyes x Lips interaction, reached significance (power=1) for all participants. At this point we can conclude that all participants exhibited non-additive effects on the MIC test. Inspection of Figure 17 (third column) shows that all MIC contrasts are positive, that is, we uniformly observed overadditivity for all participants.

Finally, the 3-way Eyes x Lips x Session interaction, that could indicate change of the MIC over the course of learning, was found to be significant for Participants 1, 4, 5 (marginally) and 6, while Participants 2 and 3 showed non-significant interaction effects. Therefore, the participants with the significant 3-way interaction could exhibit the change in architecture during learning course.

Learning by across sessions

In order to closely inspect the effect of learning for possible changes in architecture (revealed by a change in the MIC and SIC test scores), we applied similar GLM analyses using the same design, but separately for each session, for each participant (Table 2). The results supported overall analysis in Table 1: when broken into sessions, all main effects (Eyes, Lips) are significant as well as the interaction between the two (Eyes x Lips), for all participants. Exceptions to this were Participants 4 and 5. Participant 4 exhibited additivity in the first two sessions. Further inspection of the SIC curves for those sessions indicated that this participant exhibited an S-shaped SIC function which is consistent with serial exhaustive processing. Participant 5 exhibited additivity in the first block, but the main effect of Lips was not significant, which renders this interaction unusable for the SFT test, and will not be considered.

Table 2: The GLM univariate analysis was conducted on the gang-members, for different participants (SUB) across the learning sessions (in successive blocks of rows). The main effects and interaction terms are listed in the second column (*Factor*). Degrees of freedom are in the third column (*df*). The *error* row defines the degrees of freedom for the F-test error term, for that participant. Each F-test value (*F*) has two degrees of freedom: one from its corresponding row, and the other from the error row. The significance level is presented in the column *Sig.*, and the observed power for that effect is in the last column.

	Factor	df	F	Sig.	Observed Power		Factor	df	F	Sig.	Observed Power
Sub 01	Trial order	1	9.997	.002	.884	Sub 02	Trial order	1	.245	.621	.078
Session 01	Eyes	1	10.794	.001	.906	**Session 01**	Eyes	1	11.919	.001	.931
	Lips	1	103.311	.000	1.000		Lips	1	15.900	.000	.978
	Eyes x Lips	1	6.764	.010	.737		Eyes x Lips	1	9.228	.002	.858
	Error	483					Error	537			
Session 02	Trial order	1	11.009	.001	.912	**Session 02**	Trial order	1	.069	.793	.058
	Eyes	1	92.204	.000	1.000		Eyes	1	19.167	.000	.992
	Lips	1	217.179	.000	1.000		Lips	1	34.301	.000	1.000
	Eyes x Lips	1	61.101	.000	1.000		Eyes x Lips	1	9.676	.002	.874
	Error	562					Error	575			
Session 03	Trial order	1	.467	.495	.105	**Session 03**	Trial order	1	5.104	.024	.616
	Eyes	1	208.825	.000	1.000		Eyes	1	35.677	.000	1.000
	Lips	1	265.514	.000	1.000		Lips	1	57.901	.000	1.000
	Eyes x Lips	1	121.202	.000	1.000		Eyes x Lips	1	11.604	.001	.925
	Error	585					Error	579			
Session 04	Trial order	1	.849	.357	.151	**Session 04**	Trial order	1	.584	.445	.119
	Eyes	1	101.726	.000	1.000		Eyes	1	43.146	.000	1.000
	Lips	1	60.409	.000	1.000		Lips	1	15.171	.000	.973
	Eyes x Lips	1	24.198	.000	.998		Eyes x Lips	1	9.445	.002	.866
	Error	573					Error	587			
Session 05	Trial order	1	1.138	.287	.187	**Session 05**	Trial order	1	3.570	.059	.471
	Eyes	1	67.760	.000	1.000		Eyes	1	51.965	.000	1.000
	Lips	1	43.210	.000	1.000		Lips	1	52.978	.000	1.000
	Eyes x Lips	1	32.942	.000	1.000		Eyes x Lips	1	36.297	.000	1.000
	Error	584					Error	572			
Session 06	Trial order	1	5.120	.024	.618	**Session 06**	Trial order	1	.082	.775	.059
	Eyes	1	113.679	.000	1.000		Eyes	1	44.648	.000	1.000

	Lips	1	56.311	.000	1.000		Lips	1	45.369	.000	1.000
	Eyes x Lips	1	58.566	.000	1.000		Eyes x Lips	1	19.992	.000	.994
	Error	570					Error	575			
	Factor	**df**	**F**	**Sig.**	**Observed Power**		**Factor**	**df**	**F**	**Sig.**	**Observed Power**
Sub 03	Trial order	1	14.642	.000	.969	**Sub 04**	Trial order	1	14.732	.000	.969
Session 01	Eyes	1	48.063	.000	1.000	**Session 01**	Eyes	1	74.062	.000	1.000
	Lips	1	27.499	.000	.999		Lips	1	6.135	.014	.696
	Eyes x Lips	1	5.528	.019	.651		Eyes x Lips	1	.009	.926	.051
	Error	556					Error	503			
Session 02	Trial order	1	72.714	.000	1.000	**Session 02**	Trial order	1	26.856	.000	.999
	Eyes	1	59.376	.000	1.000		Eyes	1	53.657	.000	1.000
	Lips	1	54.161	.000	1.000		Lips	1	5.512	.019	.650
	Eyes x Lips	1	32.198	.000	1.000		Eyes x Lips	1	.507	.477	.110
	Error	585					Error	578			
Session 03	Trial order	1	6.771	.010	.738	**Session 03**	Trial order	1	2.269	.133	.324
	Eyes	1	30.216	.000	1.000		Eyes	1	47.260	.000	1.000
	Lips	1	36.555	.000	1.000		Lips	1	10.686	.001	.904
	Eyes x Lips	1	14.029	.000	.962		Eyes x Lips	1	16.020	.000	.979
	Error	584					Error	585			
Session 04	Trial order	1	.794	.373	.144	**Session 04**	Trial order	1	.467	.495	.105
	Eyes	1	59.436	.000	1.000		Eyes	1	88.685	.000	1.000
	Lips	1	49.918	.000	1.000		Lips	1	18.593	.000	.990
	Eyes x Lips	1	25.413	.000	.999		Eyes x Lips	1	7.169	.008	.762
	Error	582					Error	587			
Session 05	Trial order	1	2.337	.127	.333	**Session 05**	Trial order	1	.041	.839	.055
	Eyes	1	78.106	.000	1.000		Eyes	1	87.864	.000	1.000
	Lips	1	83.123	.000	1.000		Lips	1	16.863	.000	.984
	Eyes x Lips	1	26.577	.000	.999		Eyes x Lips	1	9.754	.002	.877
	Error	574					Error	587			
Session 06	Trial order	1	.634	.426	.125	**Session 06**	Trial order	1	.195	.659	.073
	Eyes	1	120.220	.000	1.000		Eyes	1	81.566	.000	1.000
	Lips	1	56.914	.000	1.000		Lips	1	23.671	.000	.998
	Eyes x Lips	1	34.858	.000	1.000		Eyes x Lips	1	23.974	.000	.998
	Error	581					Error	585			

	Factor	df	F	Sig.	Observed Power		Factor	df	F	Sig.	Observed Power

Sub 05	Trial order	1	5.104	.024	.616	**Sub 06**	Trial order	1	.580	.447	.118
Session 01	Eyes	1	41.723	.000	1.000	**Session 01**	Eyes	1	43.785	.000	1.000
	Lips	1	37.944	.000	1.000		Lips	1	2.314	.129	.330
	Eyes x Lips	1	6.161	.013	.698		Eyes x Lips	1	1.048	.306	.176
	Error	543					Error	542			
Session 02	Trial order	1	1.825	.177	.271	**Session 02**	Trial order	1	3.438	.064	.457
	Eyes	1	114.947	.000	1.000		Eyes	1	117.902	.000	1.000
	Lips	1	93.753	.000	1.000		Lips	1	52.711	.000	1.000
	Eyes x Lips	1	44.274	.000	1.000		Eyes x Lips	1	36.848	.000	1.000
	Error	575					Error	568			
Session 03	Trial order	1	2.077	.150	.301	**Session 03**	Trial order	1	.683	.409	.131
	Eyes	1	128.824	.000	1.000		Eyes	1	82.785	.000	1.000
	Lips	1	76.313	.000	1.000		Lips	1	59.334	.000	1.000
	Eyes x Lips	1	38.951	.000	1.000		Eyes x Lips	1	14.187	.000	.964
	Error	575					Error	567			
Session 04	Trial order	1	6.576	.011	.726	**Session 04**	Trial order	1	9.791	.002	.878
	Eyes	1	119.696	.000	1.000		Eyes	1	122.621	.000	1.000
	Lips	1	39.764	.000	1.000		Lips	1	112.211	.000	1.000
	Eyes x Lips	1	23.838	.000	.998		Eyes x Lips	1	65.102	.000	1.000
	Error	582					Error	580			
Session 05	Trial order	1	4.425	.036	.556	**Session 05**	Trial order	1	11.411	.001	.921
	Eyes	1	107.637	.000	1.000		Eyes	1	135.056	.000	1.000
	Lips	1	48.106	.000	1.000		Lips	1	84.418	.000	1.000
	Eyes x Lips	1	32.534	.000	1.000		Eyes x Lips	1	52.414	.000	1.000
	Error	590					Error	581			
Session 06	Trial order	1	.692	.406	.132	**Session 06**	Trial order	1	1.607	.205	.244
	Eyes	1	135.406	.000	1.000		Eyes	1	48.841	.000	1.000
	Lips	1	41.245	.000	1.000		Lips	1	29.918	.000	1.000
	Eyes x Lips	1	29.844	.000	1.000		Eyes x Lips	1	17.283	.000	.986
	Error	595					Error	576			

Learning sessions and MIC

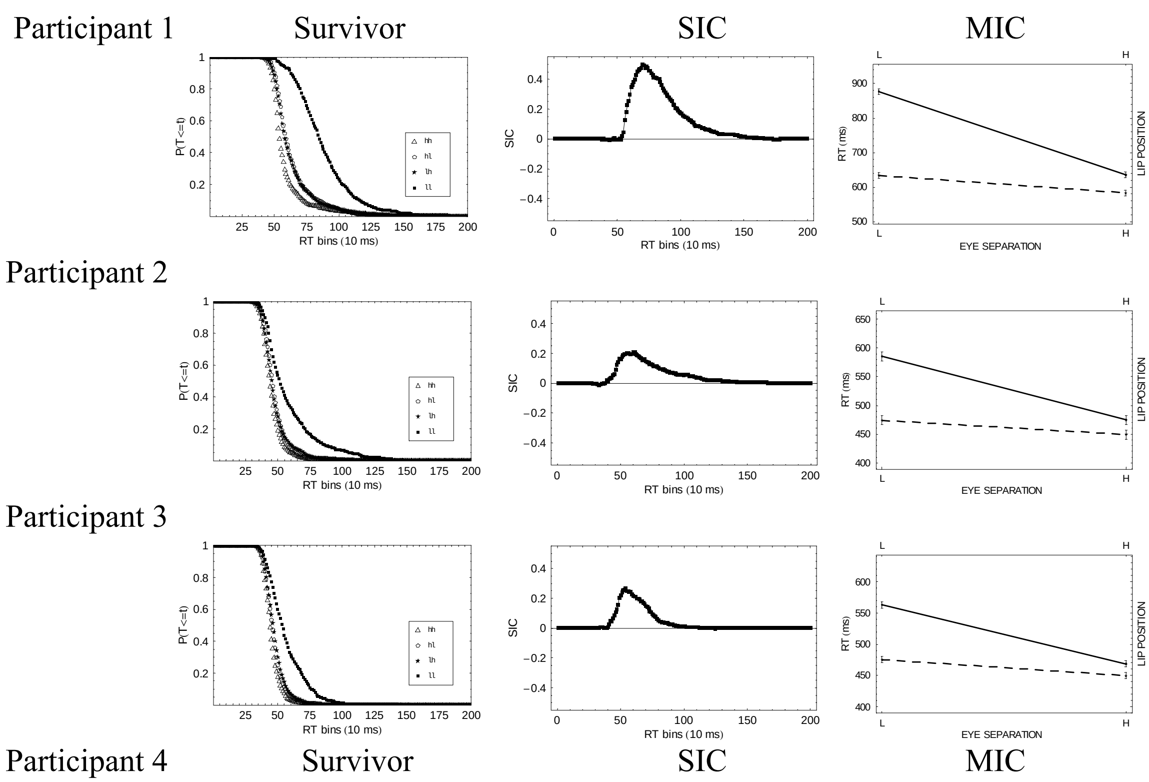
We investigated the relationship between learning block and the change of the MIC. In the GLM analyses some participants showed significant change of MIC score as leaning progressed (as a function of learning sessions). This finding suggests that the MIC value should change as a function of the sessions. In order to test this, we ran a multiple regression analysis using subjects' MIC scores as the dependent variable, while the session number, the mean RT, and their interaction were independent variables. The proportion of explained variability of the MIC change was $R^2=0.58$, $F(2,35)=22.83$, $p<0.01$, with both the block and mean RT contributing significantly. Given the obtained proportion of unexplained variability, the change in the MIC values can not be completely attributed to learning. When averaged over participants, the MIC value showed increase only from the first learning session, $MIC_{block1}=41ms$, to around 100ms value for the other sessions. On average, this increase indicates strong MIC value change between the first and the rest of sessions only.

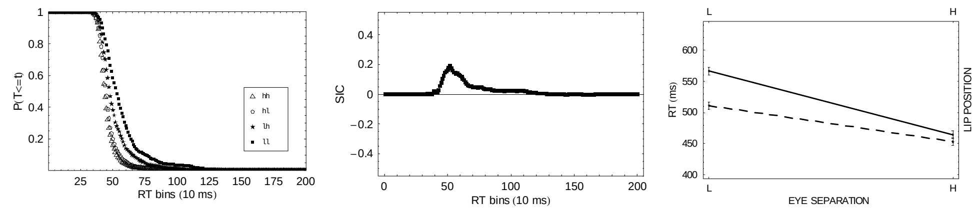
Mean and Survivor Interaction Contrast Functions

Both the MIC and SIC are powerful tools for investigating the organization of mental processes. In fact, the SIC function provides more information than the MIC test by allowing more inferential power. When integrated over time, the SIC function returns the exact MIC score. It is obvious then, that different shapes of SIC function could predict the same MIC value. Therefore, in our analyses we will graph both functions.

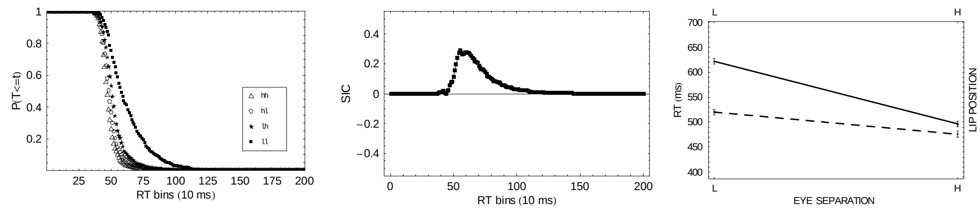
In Figure 17 we show the MIC results for different participants, paired with their corresponding SIC functions. Note that both tests are calculated for RTs collapsed across learning sessions. We also provide an additional corresponding plot of the survivor functions used to calculate the SIC function (Figure 17, left column). Survivor functions are nicely ordered for all participants (LL>LH≈HL>HH), which implies a corresponding ordering of the means. Notably, the results for both the MIC and SIC reveal parallel architecture with minimum time stopping rule. The results are uniform over all participants.

Figure 17: The SFT tests results for the OR condition, for gang-member faces, for all participants. The results are based on all learning sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.

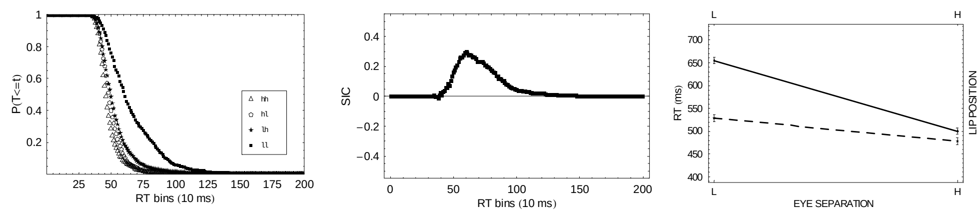




Participant 5



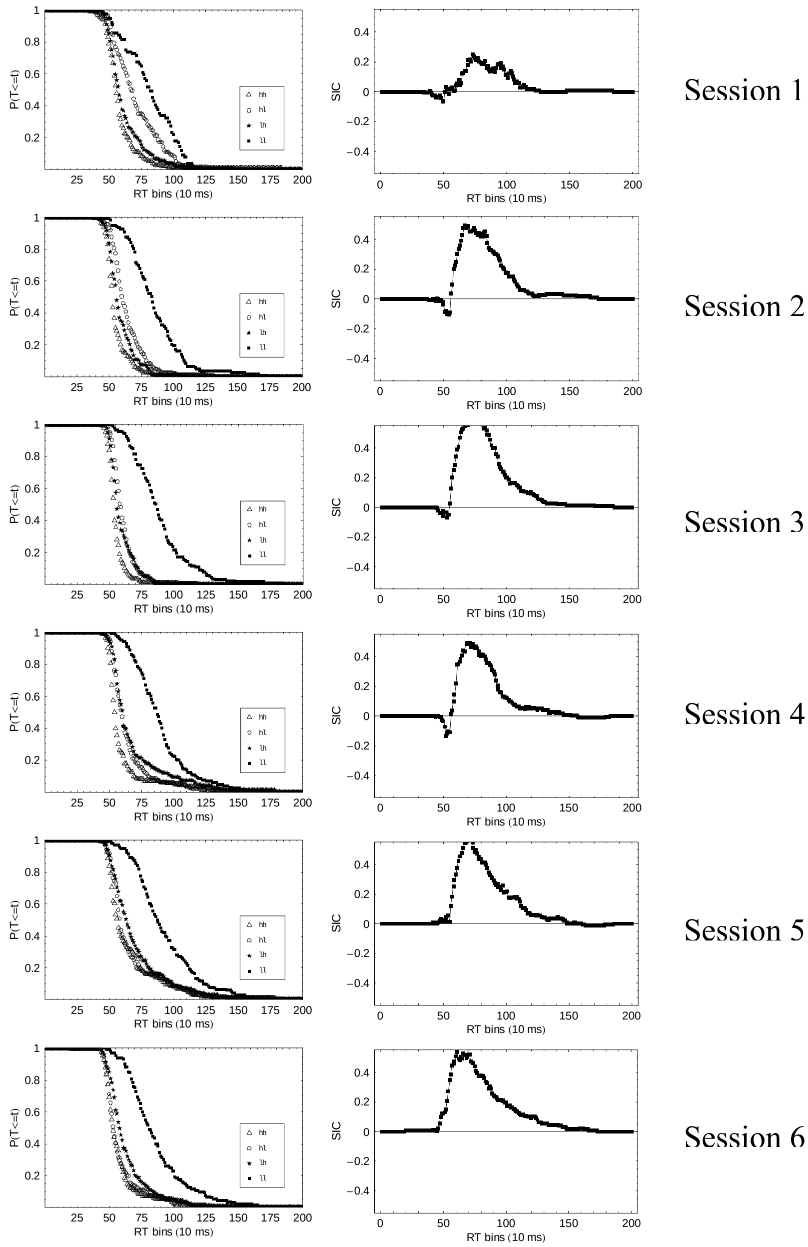
Participant 6



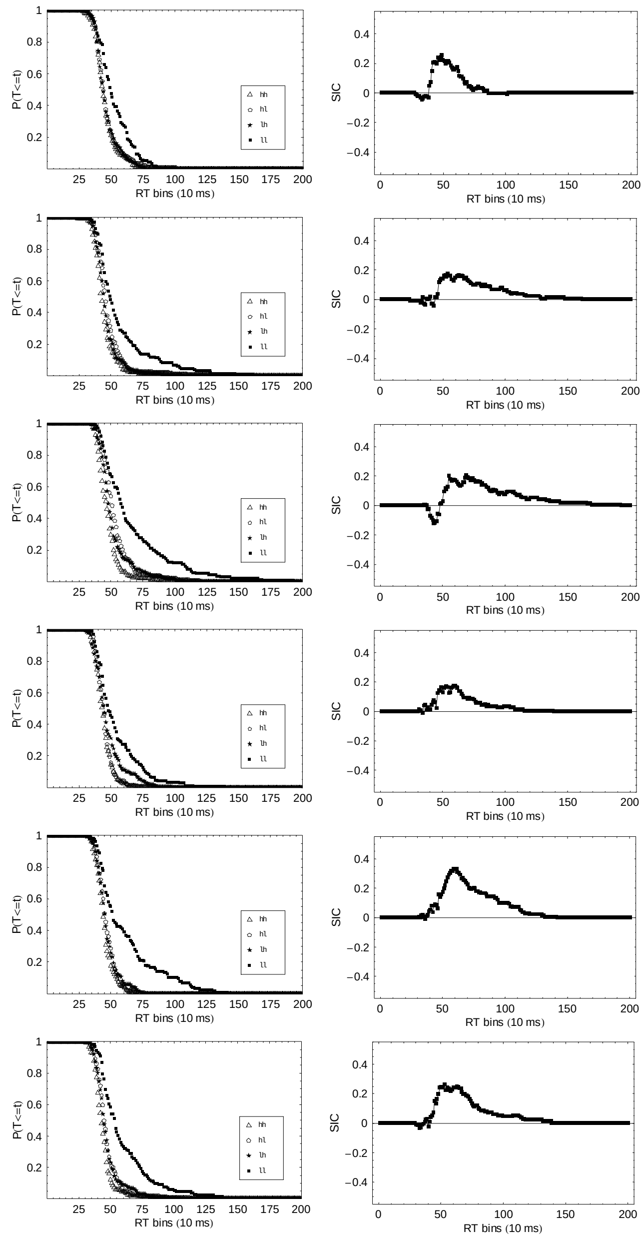
In the Figure 17 B we separated the both tests into different learning sessions. In accordance with previous analyses based on GLM tests, and following overall finding of overadditivity on both the MIC and SIC levels, we revealed the same signature of parallel minimum time processing architecture in almost all conditions, for different participants.

Figure 17 B: The SFT tests results for the OR condition, for gang-member faces, for all participants. The results are broken down for each participant across the learning sessions. The first column depicts the ordering of joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the right column. The learning sessions are presented in rows.

Participant 1 *Survivor* *SIC*



Participant 2 *Survivor* *SIC*



Session 1

Session 2

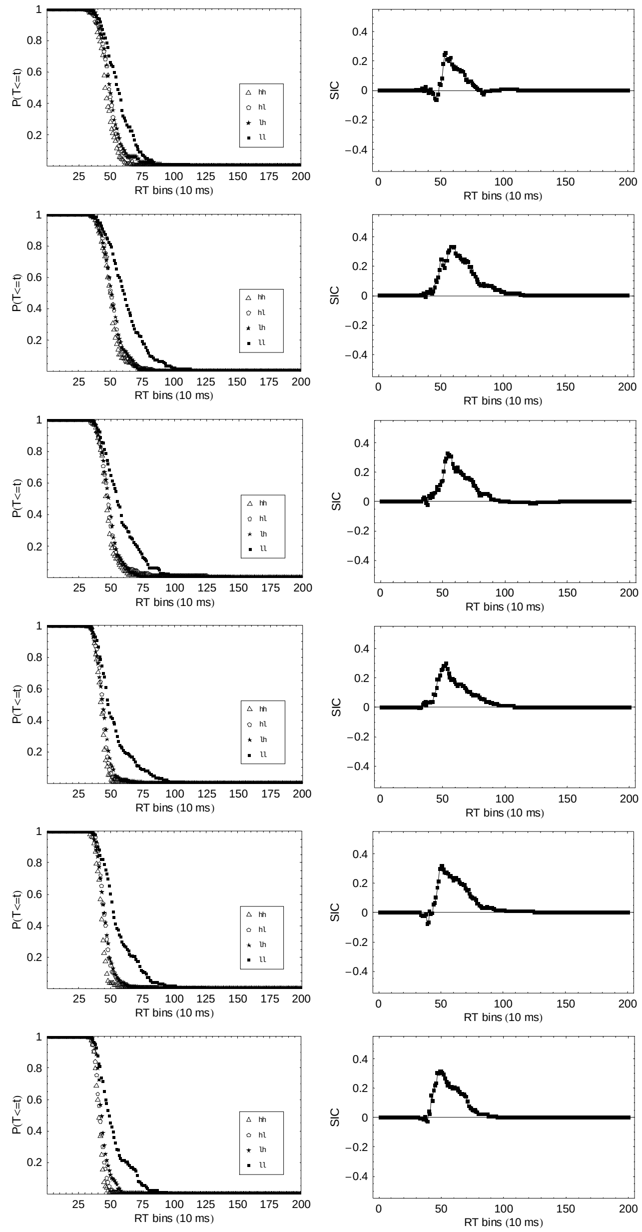
Session 3

Session 4

Session 5

Session 6

Participant 3 *Survivor* *SIC*



Session 1

Session 2

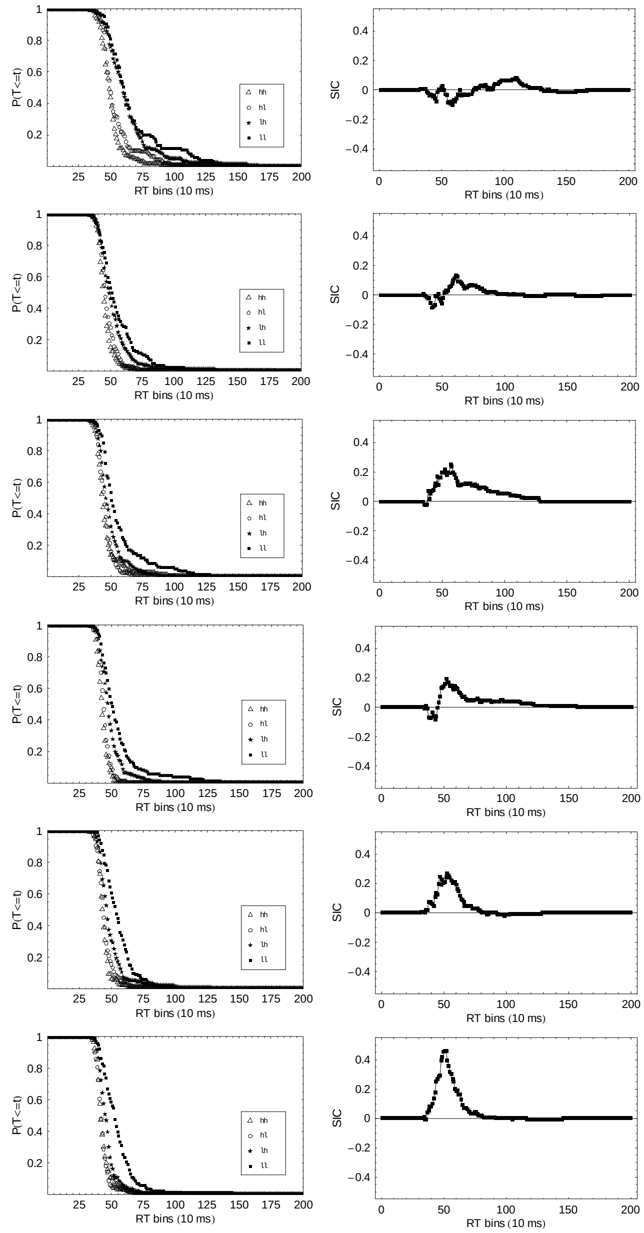
Session 3

Session 4

Session 5

Session 6

Participant 4 *Survivor* *SIC*



Session 1

Session 2

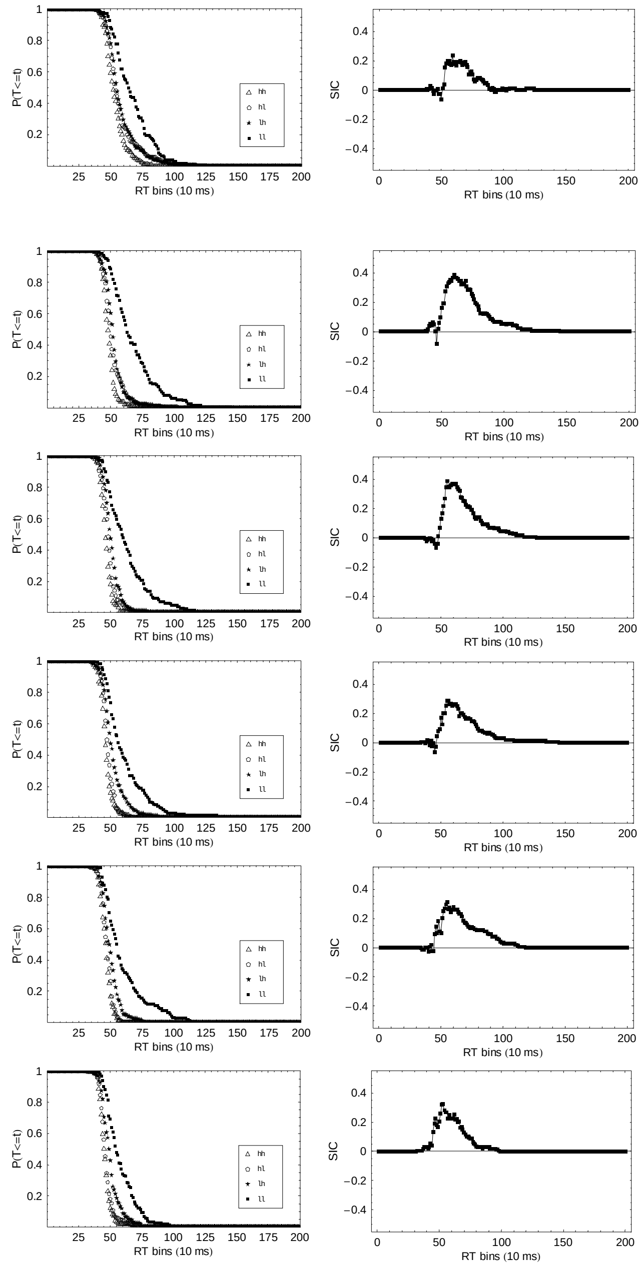
Session 3

Session 4

Session 5

Session 6

Participant 5 *Survivor* *SIC*



Session 1

Session 2

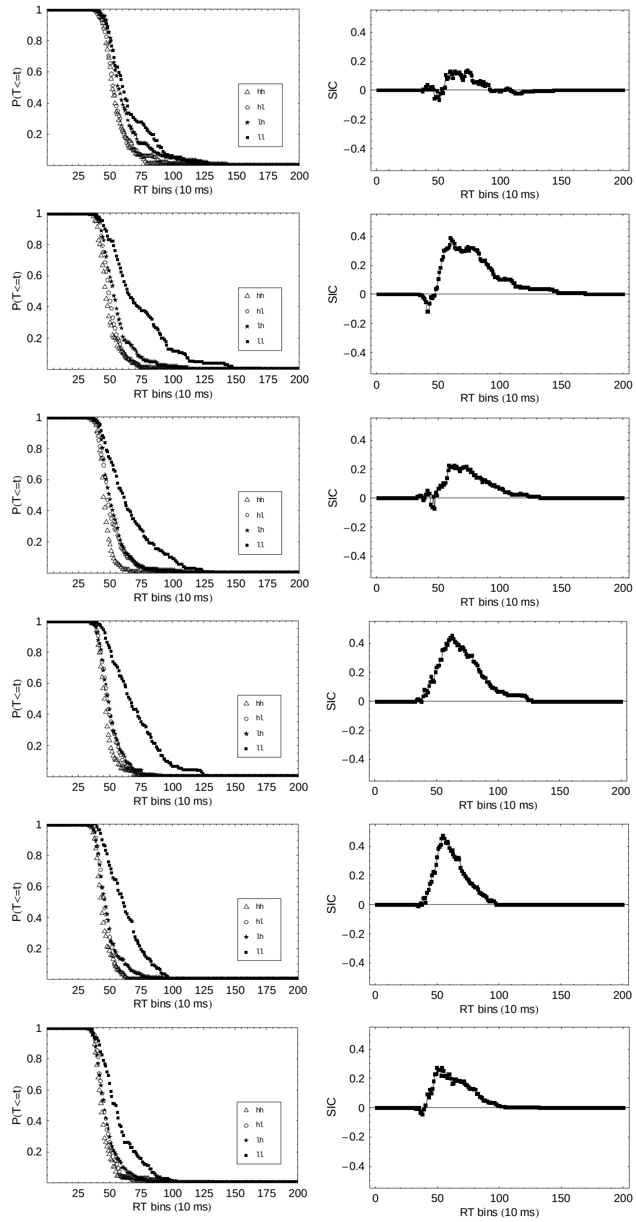
Session 3

Session 4

Session 5

Session 6

Participant 6 *Survivor* *SIC*



Session 1

Session 2

Session 3

Session 4

Session 5

Session 6

Again, the exception to this is Participant 4, who exhibited additivity on the MIC testing using GLM, and S-shaped SIC functions, in the first two learning sessions, which is strong support for serial exhaustive strategy in early learning.

It is interesting to note that some of the calculated SIC functions posses the small negative blip signature for fast reaction times. However appearance of the negative blip is not consistent over the learning sessions, and usually can be observed in the midst of the learning sessions, rather than in the end, where it would be expected. This expectation is based on the fact that if the negative blip is a signature of coactive processing, then it would be more likely that coactive processing develops later in the learning process. However, this is not evident in our results (Figure 17 B).

Difference between blocked and mixed whole-face conditions

We adopted a null hypothesis which states that there is no difference on the mean RT level between blocked and mixed conditions when whole-faces are categorized. In the blocked condition, only whole-faces were presented on each trial. In the mixed condition, both whole-faces and masked-faces were combined together in the experimental block, and were randomly presented in the experiment. No mean RT difference suggests that the two forms of representation, or processes (whole- and masked-faces) are independent and not influenced directly or indirectly by some third agent. Direct independence is in agreement with the idea that the two processes do not share any common mechanism or representational format. If they have a common structure/representation then processing both types of stimuli in the mixed condition will produce a difference in processing relative to the condition in which only one type is presented. The indirect form of dependence hypothesis assumes that both types are processed independently but could be connected to a common agent; for example, they may share processing resources. For example, if both types (whole and masked) require the same amount of capacity, then mixed presentation of both will place more demands on the system's capacity and that can slow down processing of both, in contrast to when only one is presented.

The idea that both representations (whole- and part-faces) are utilized in the system during face encoding has long history (see the recently published strong test for dual processing by Ingvalson & Wenger, 2005).

In order to test the null hypothesis, we ran a paired sampled t-test, on the mean RT difference between processing time for whole faces in the blocked condition (the SFT test) of the learning session and processing time for whole faces in the mixed condition (from capacity test part) (see Figure 9, the 2nd part). We found that whole-faces are processed faster in the blocked condition, than in the mixed condition: $t(5)=-9.58$, $M_{blocked}=536$ms, SD=71ms, and $M_{mixed}=619$ms, SD=70ms, and this finding is consistent over all participants when ran separate analyses. We could conclude that somehow the processing of whole faces in the two conditions is influenced by the processing of part-based (masked) information. The major consequence of this finding is in the calculation of the capacity functions in the next part. Since there is significant difference of processing of whole faces depending on the presence of a part face context in the experiment, we must utilize the calculation of two additional capacity coefficient functions, each of which uses a different whole face (taken from the blocked and mixed conditions).

<u>Capacity coefficient Functions</u>

We calculated 4 different coefficient capacity functions for each participant. In Figures 18 to 25, we present the calculated capacity coefficient functions, along with bootstraped 90% percentile confidence intervals (for utilization of bootstrapping in statistical inference see (Efron & Tibshirani, 1993; Zandt, 2002)), for both gang-member faces (pooled together into one condition) and the hero-face. So for each participant, we calculated a total of 8 capacity functions, for each learning session. Confidence intervals delimit the area in which the CCF would fall most likely in, when the same experiment is repeated. We are generally interested in any violations of the capacity coefficient value equal to one. Since the shape of the calculated integrated hazard function depends on sampling properties, it is possible that apparent violations of the calculated CCF are evident but not certain, in a probabilistic way. That is, under some circumstances, the number of sampled observations may be insufficient to establish an unbiased estimated of the population capacity function. A sufficient sample size would be reasonably large, with up to several thousands of observations needed to achieve stable behavior. So, in order to reduce error of inference, we calculated 90% confidence intervals around our capacity coefficient functions, describing the area of high confidence of our statistical inference.

We assume that in the case of super capacity (that is, $C(t)>1$) both bounds of the confidence intervals should swing above the value of one. In the case of unlimited capacity, $C(t)=1$ should lie in between the two confidence intervals; in other words, the upper interval would be above one, and the lower interval would be below one. Finally, in the case of limited capacity, both confidence interval bounds should fall below the value of one.

It is also possible that at different points in the time scale, the estimated confidence intervals are a combination of cases described above. For example, at some point in time both bounds might lie above one, and at another time the bounds might include one. Then, following the definition of the capacity coefficient function and its relation to real a system property (Townsend & Nozawa, 1995) we could assume that at some point, the system is of super capacity, but is of unlimited capacity for a later time interval.

In Figures 18-25 we plot CCFs (4) for each participant (6), for both the gang member and hero faces. We grouped the CCFs with respect to the block conditions (blocked whole faces vs. mixed whole faces with masked faces), for each type of face.

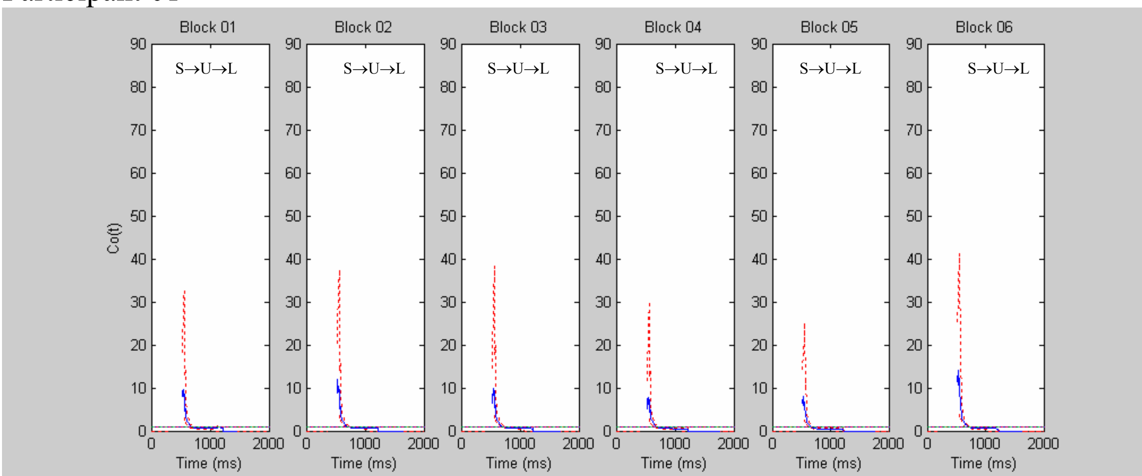
We also used scales on the y-axis that are common for all figures in the experiment, so the reader can see the changes of overall magnitude as a function of learning sessions.

On each figure we denote by a black bold line the calculated CCF; with red dotted lines we present the bootstrapped 90% confidence intervals. The C(t)=1 reference is denoted by horizontal line. Note that a violation of this value bound in any direction (super capacity or limited capacity) will show, at some time violation, by both bootstrapped confidence interval bounds. That is, both bounds will swing above or below the C(t)=1 value.

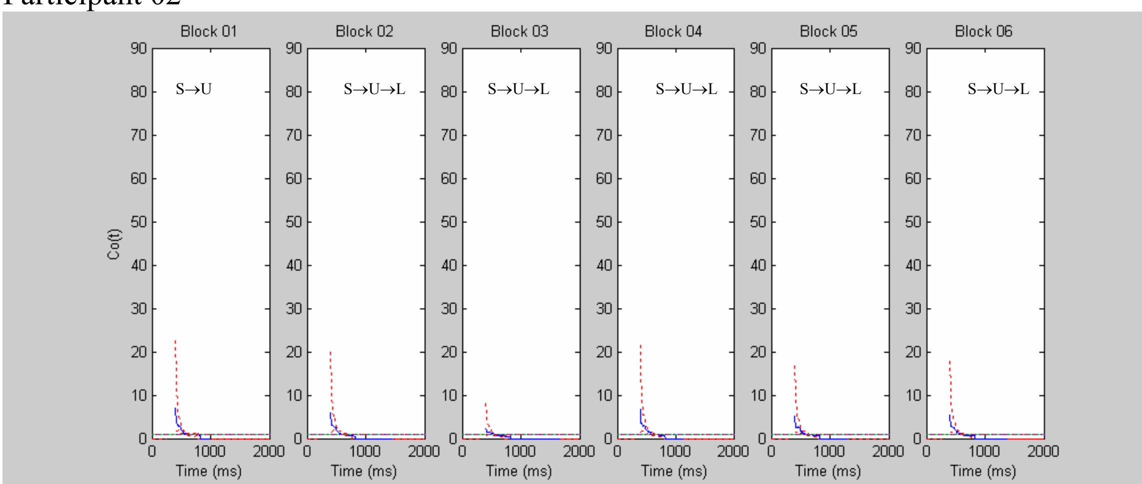
In some figures it could be difficult to verify by eye confidence interval violations of C(t) =1. Although this could be solved by rescaling of the y-axis, then a reader will miss information of magnitude across sessions. Again, we scale the y-axis such that all CCFs for one logical conditions (OR or AND) are comparable, given the highest observed CCF. In order to aid a reader's understanding of Figures 18-25, we added statistical conclusion concerning violations in each small figure: if CCF function was super, unlimited or limited capacity then we used uppercase letters S, U and L, respectively. If a transition was observed, that is if for some time CCF was super capacity and then unlimited capacity for remaining time, we used an arrow symbol. So the notation S→U→L means that the CCF exhibited all capacity states over the course of time. Also note a decreasing order of capacity is usually preserved given that CCF for OR processing typically exhibits a reduction in magnitude over time (Townsend & Wenger, 2004).

Figure 18: The **absolute learning whole-blocked CCFs**, for the **gang members**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

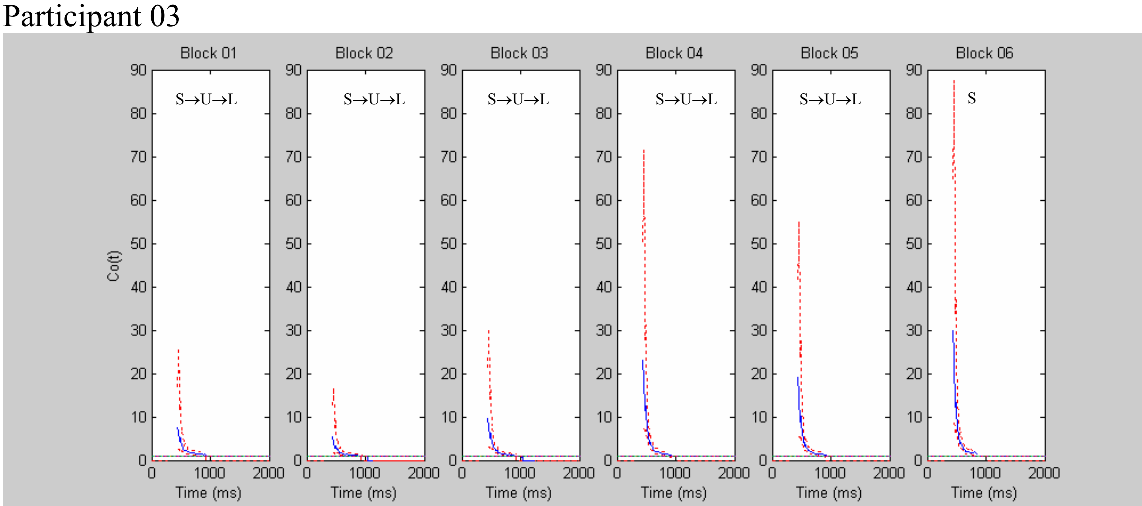
Participant 01



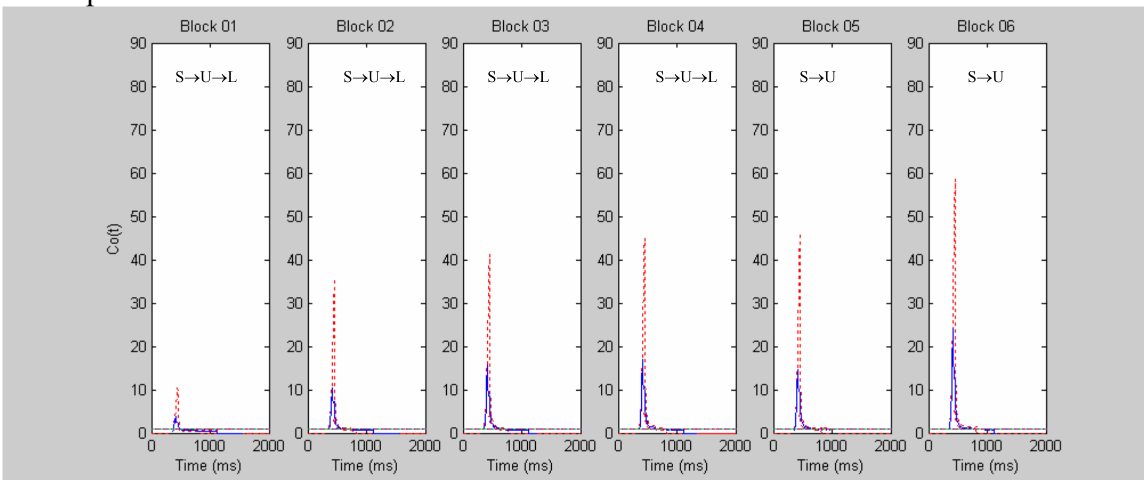
Participant 02



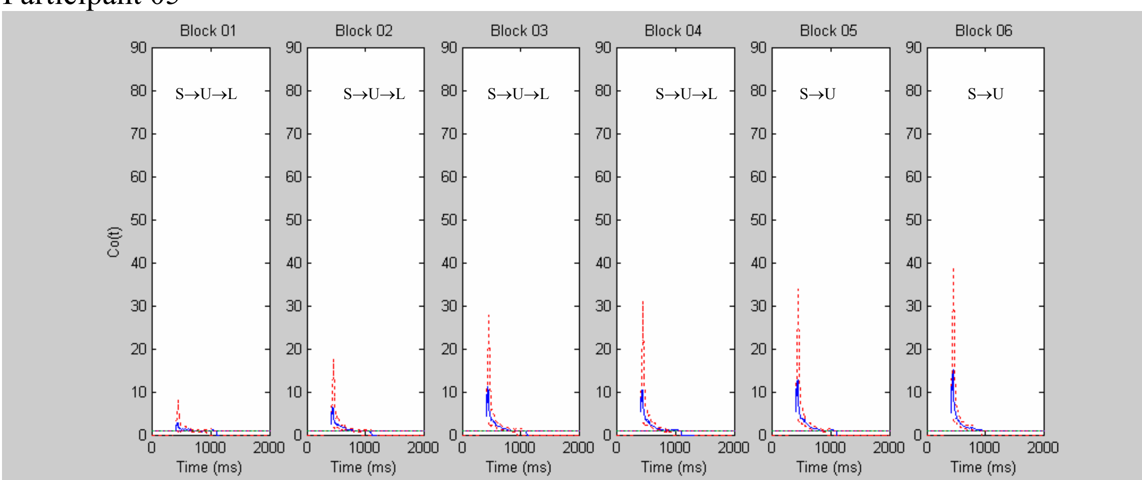
Participant 03



Participant 04



Participant 05



Participant 06

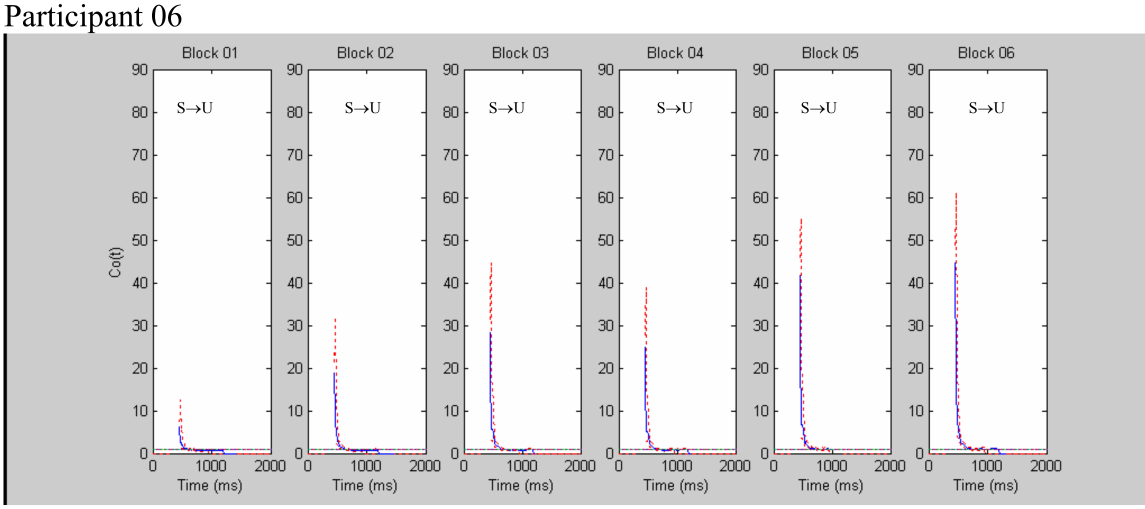
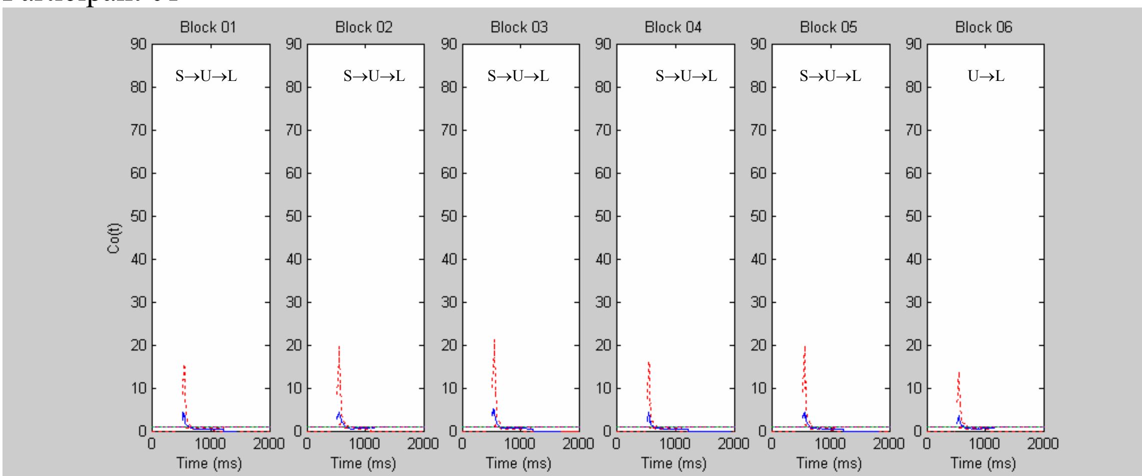
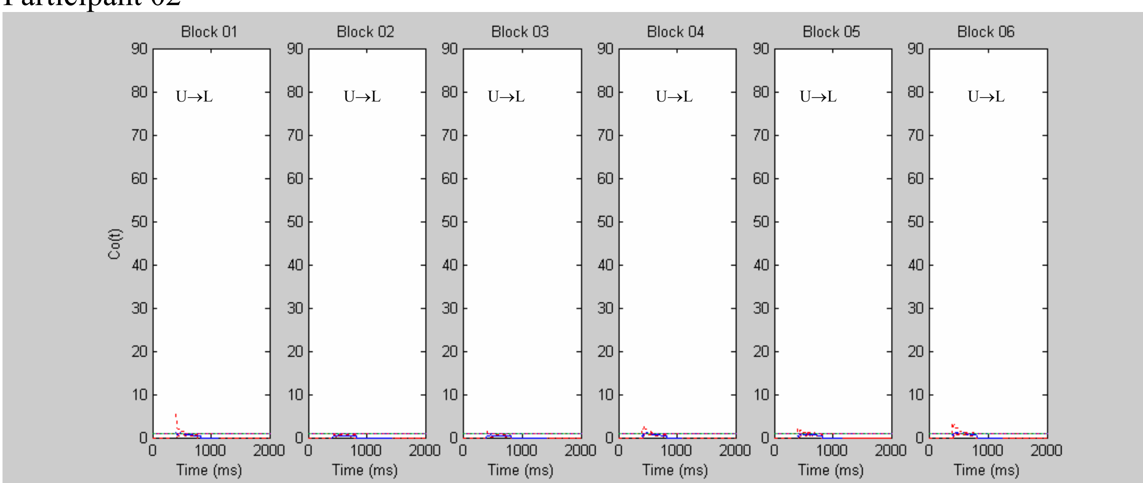


Figure 19: The **Absolute learning whole-mixed CCFs**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

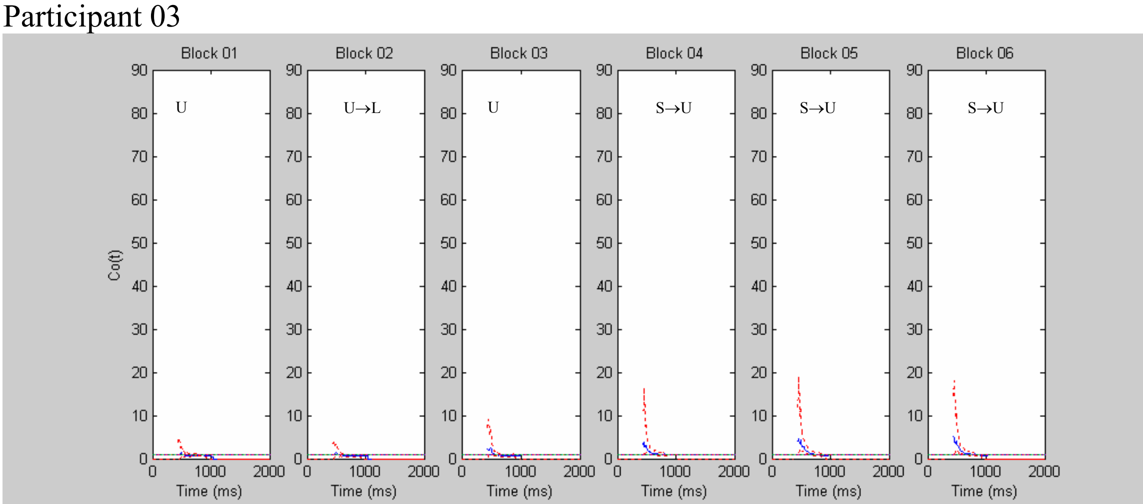
Participant 01



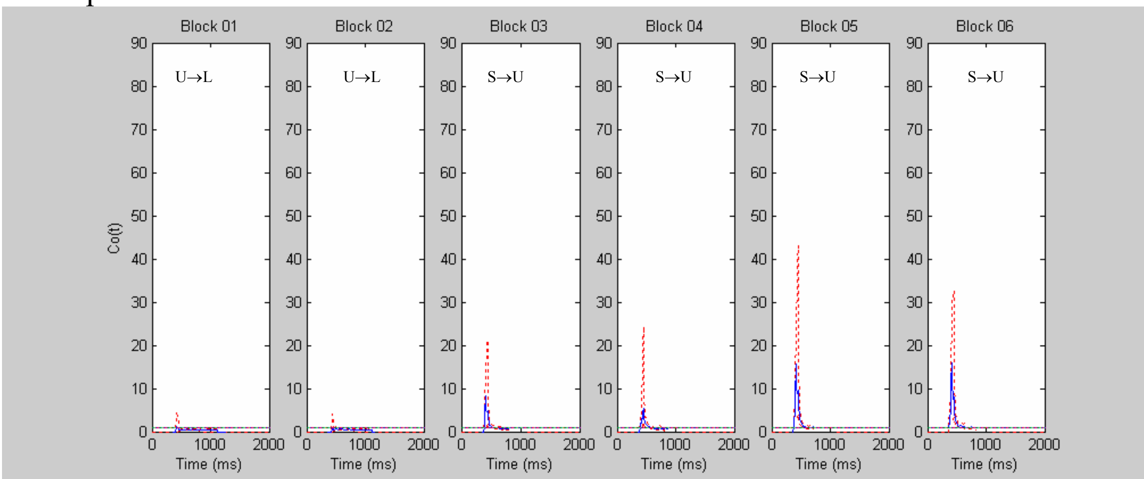
Participant 02



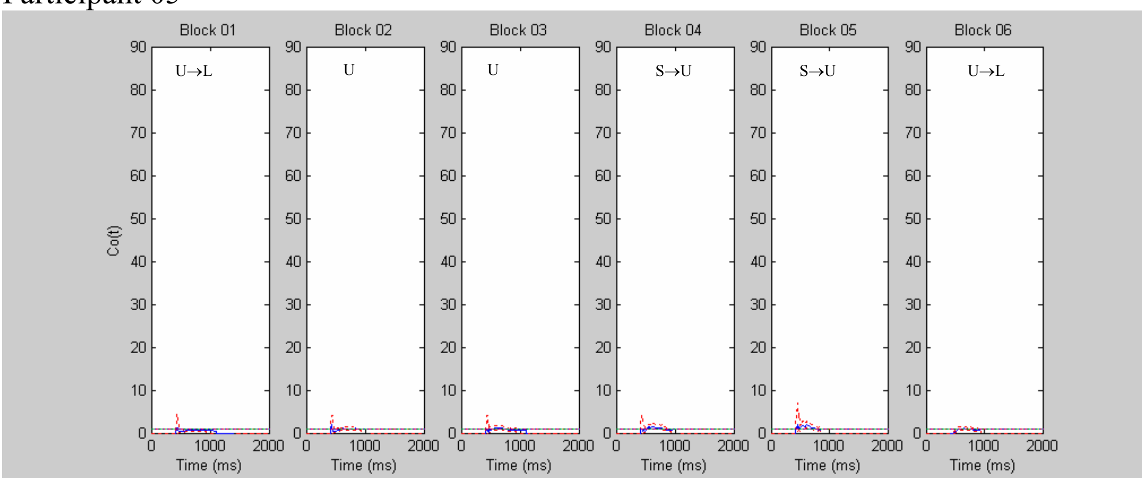
Participant 03



Participant 04



Participant 05



Participant 06

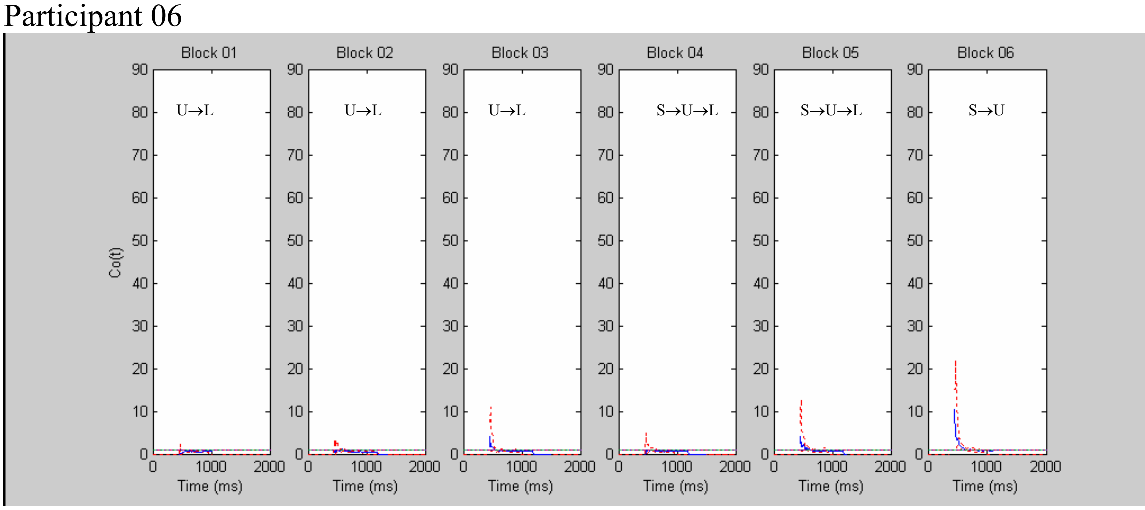
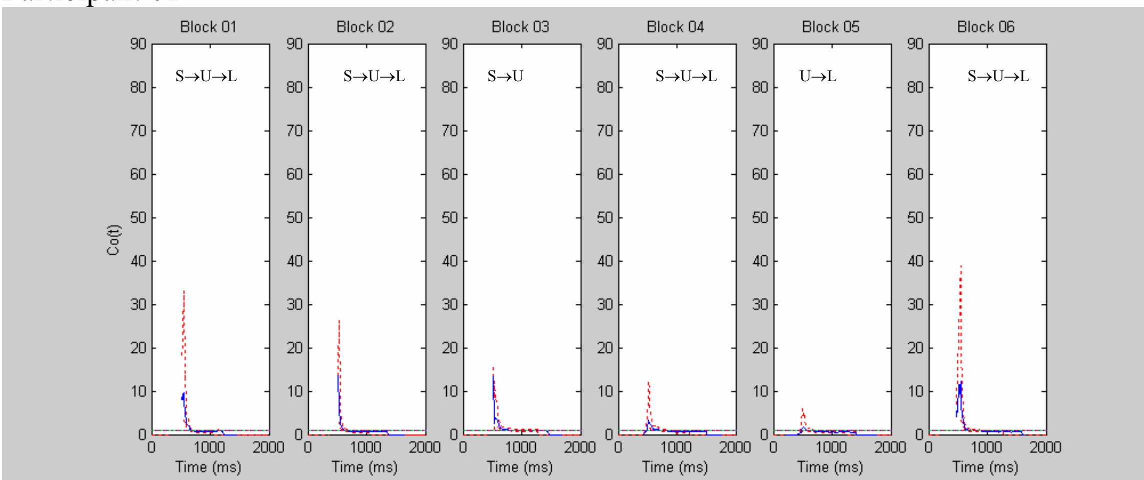
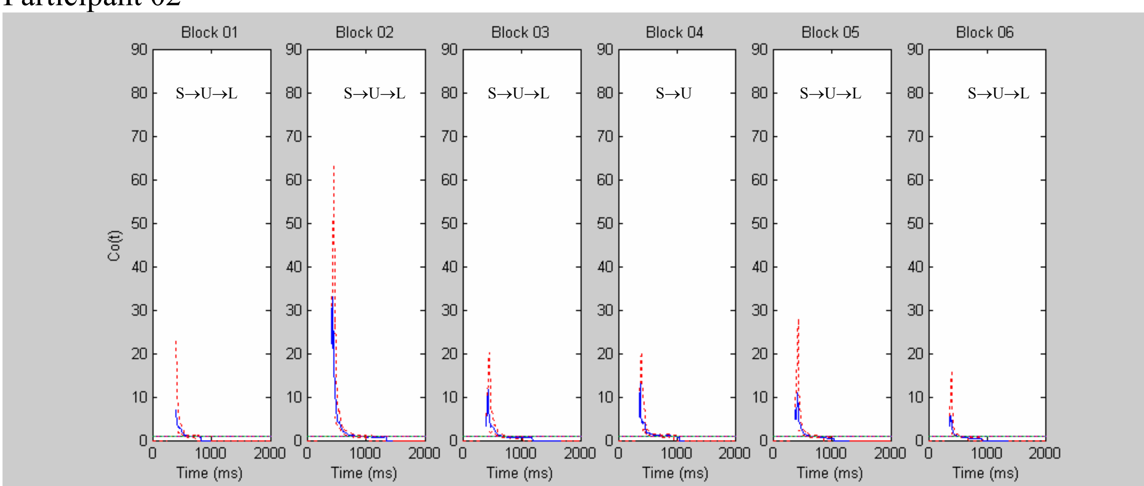


Figure 20: The **Relative learning whole-blocked CCFs**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

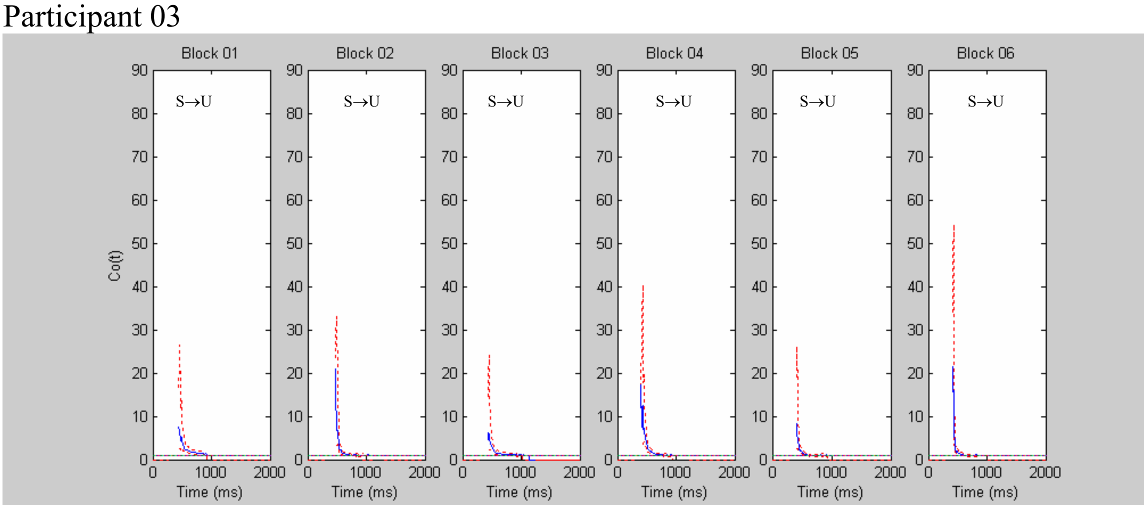
Participant 01



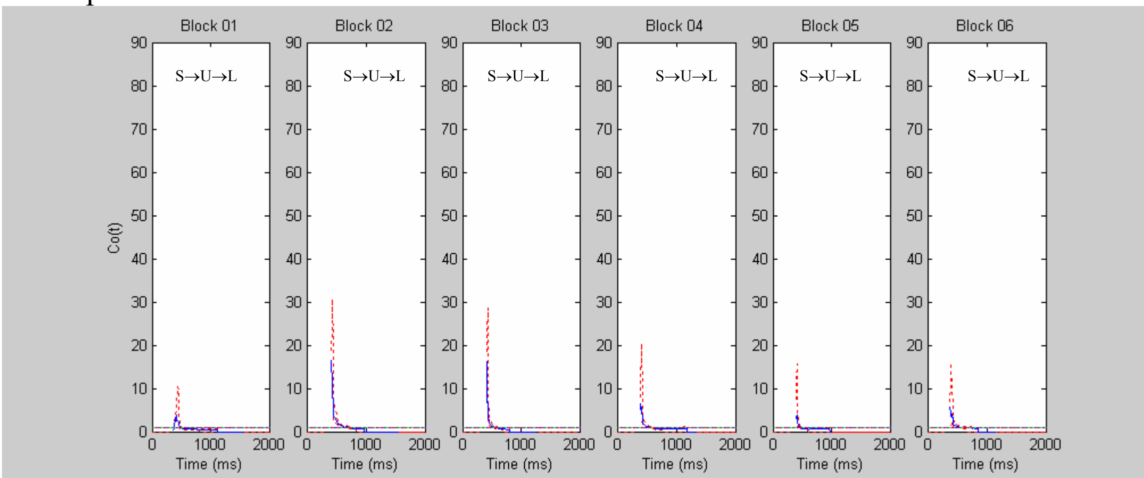
Participant 02



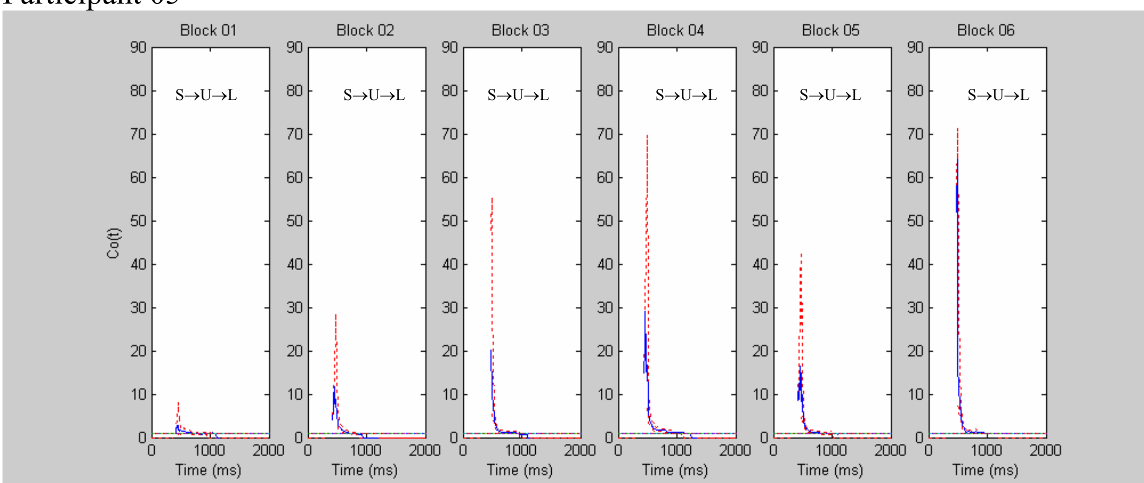
Participant 03



Participant 04



Participant 05



Participant 06

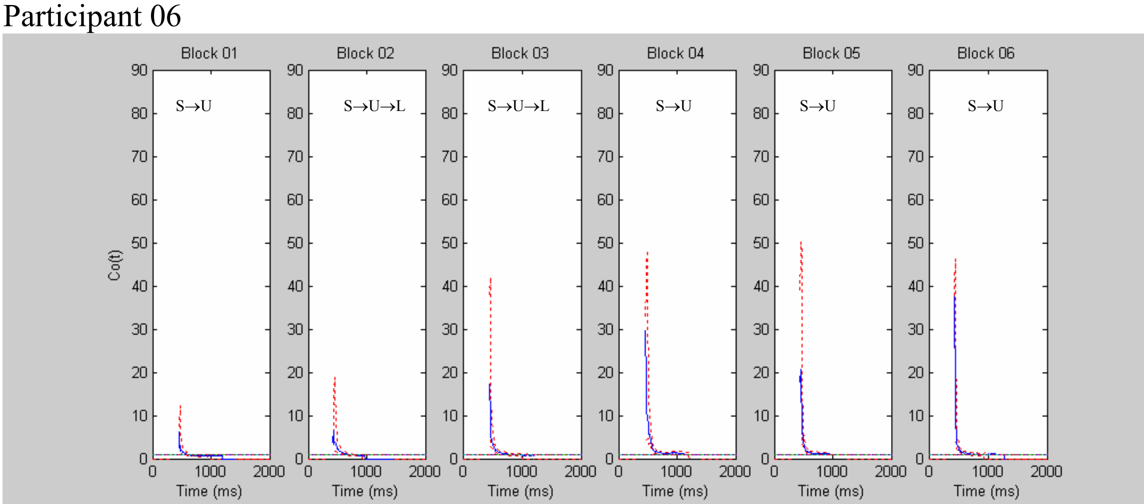
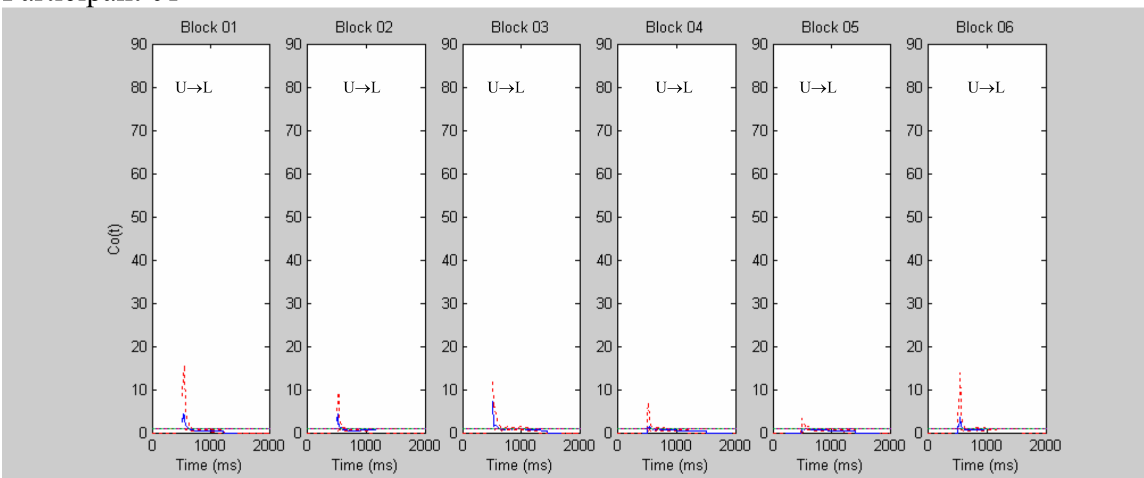
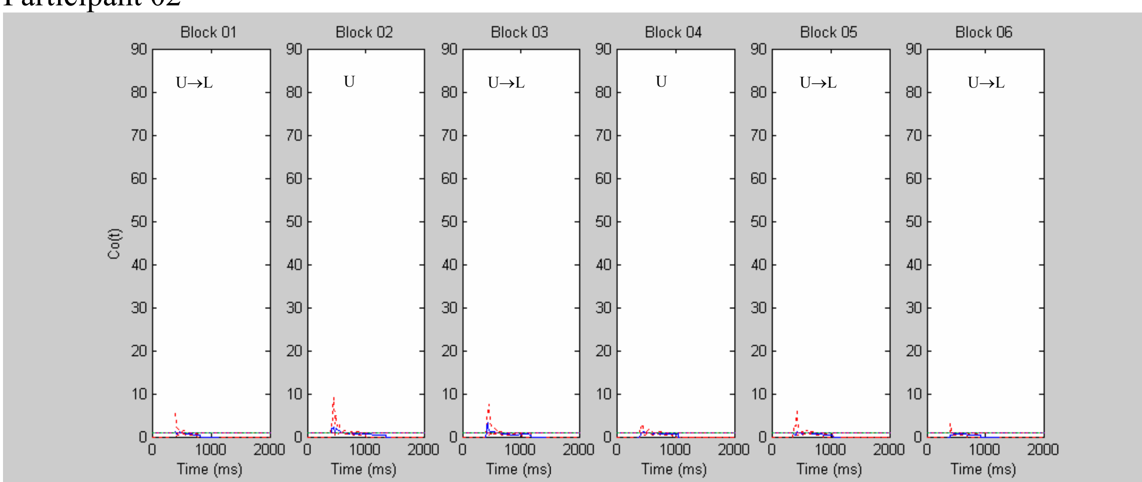


Figure 21: The **Relative learning whole-mixed CCFs**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

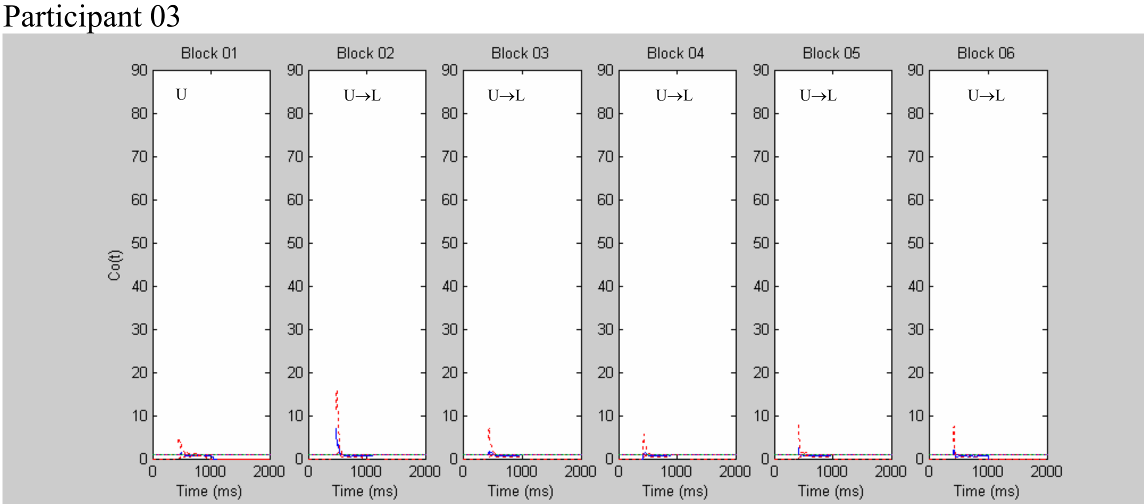
Participant 01



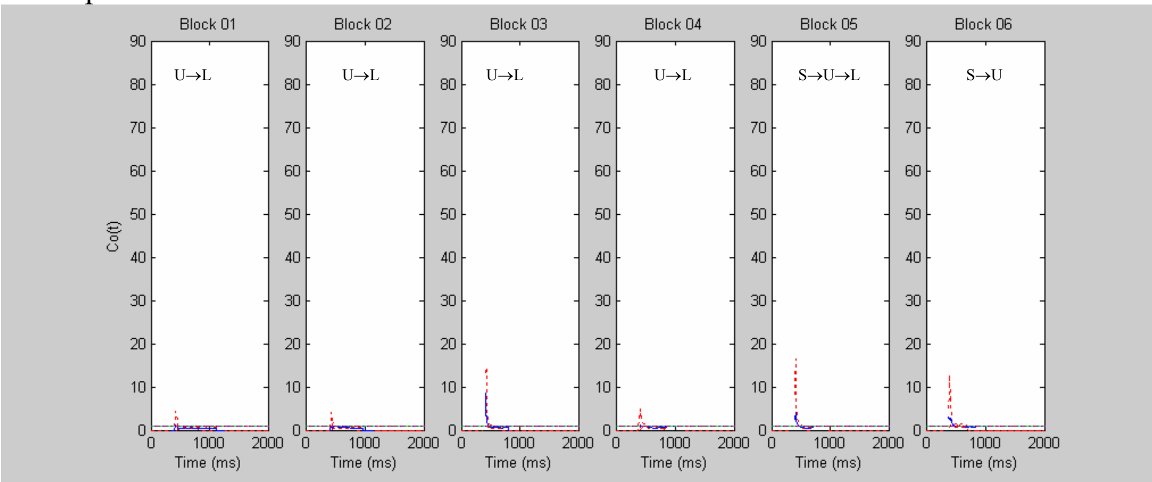
Participant 02



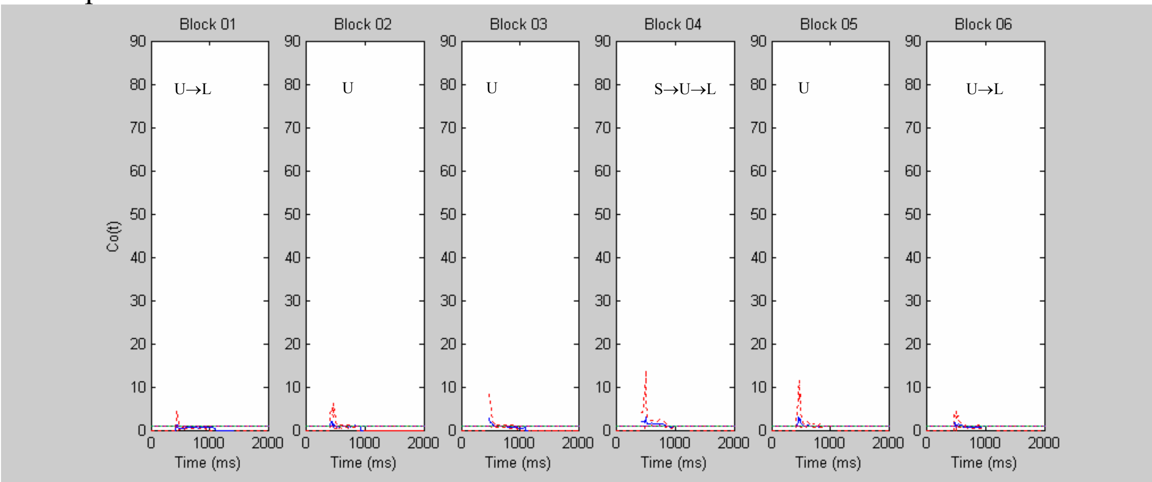
Participant 03



Participant 04



Participant 05



Participant 06

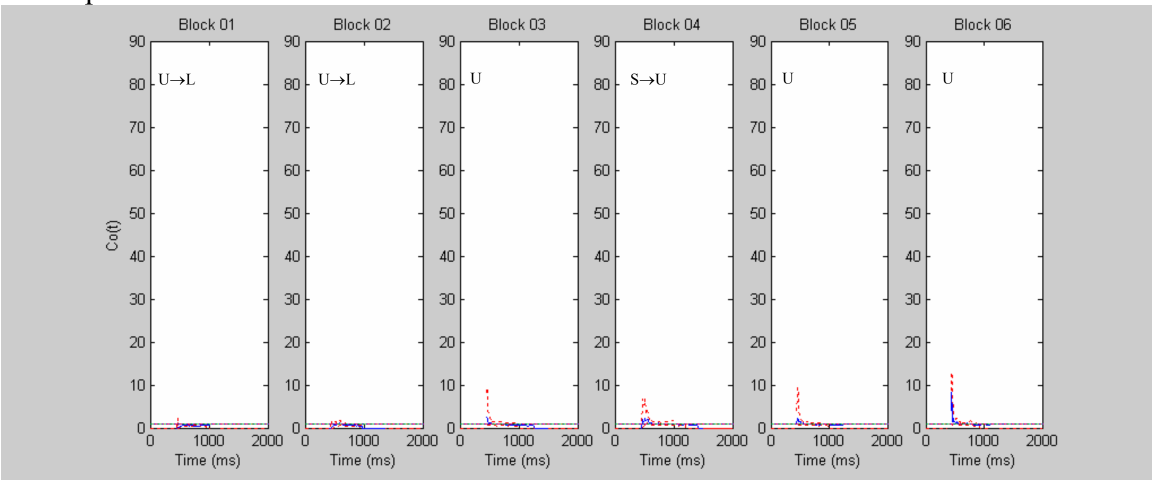
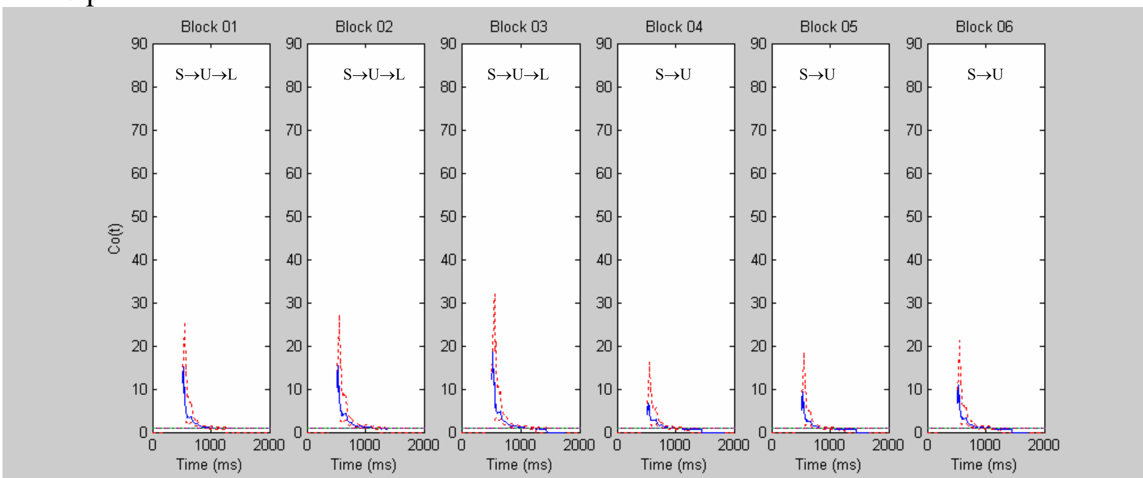
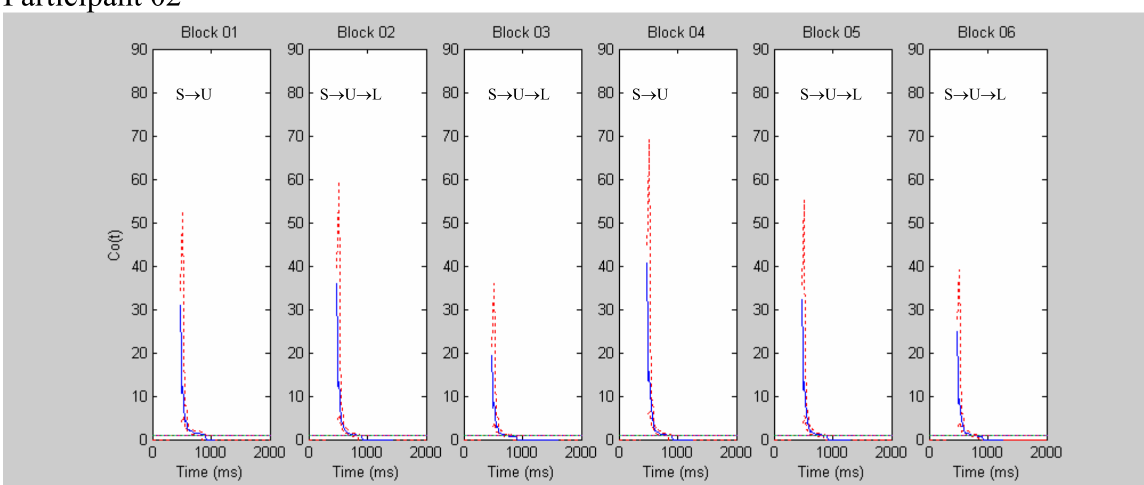


Figure 22: The **Absolute learning whole-blocked CCFs**, for the **Hero** across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

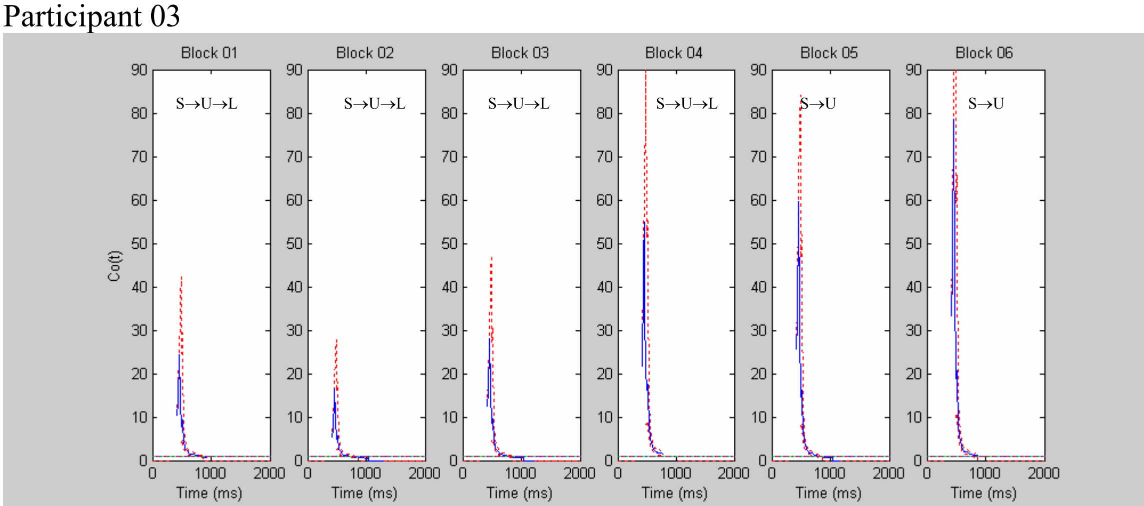
Participant 01



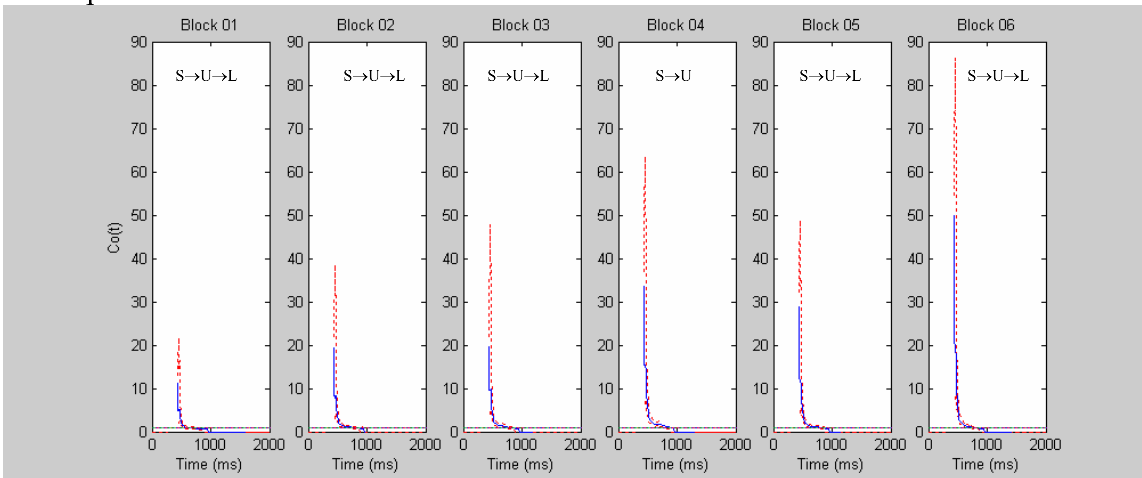
Participant 02



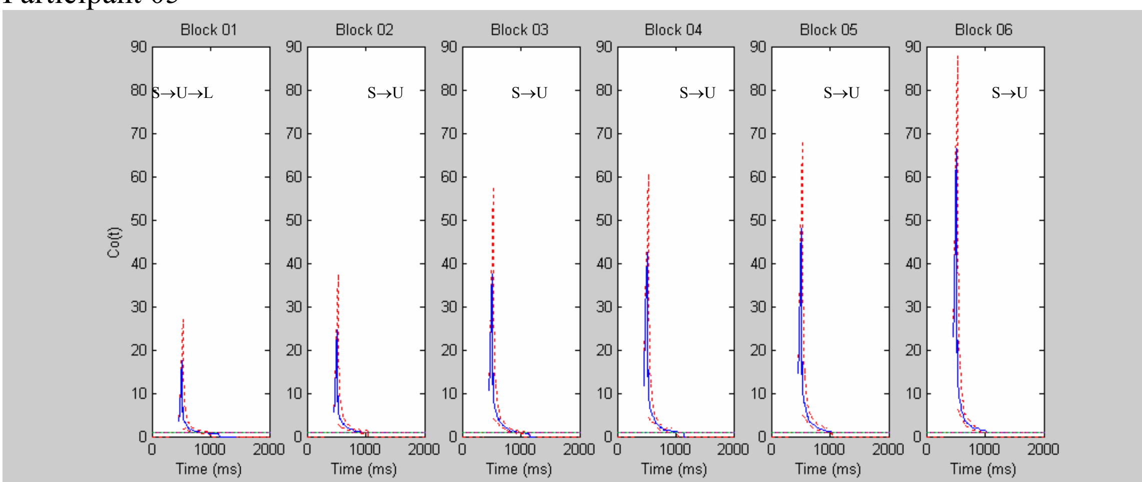
Participant 03



Participant 04



Participant 05



Participant 06

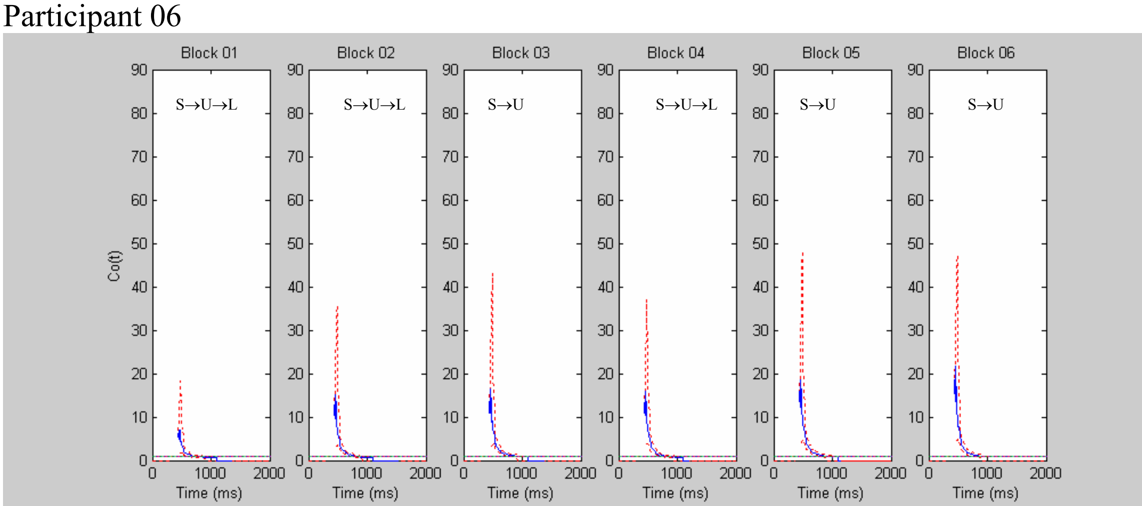
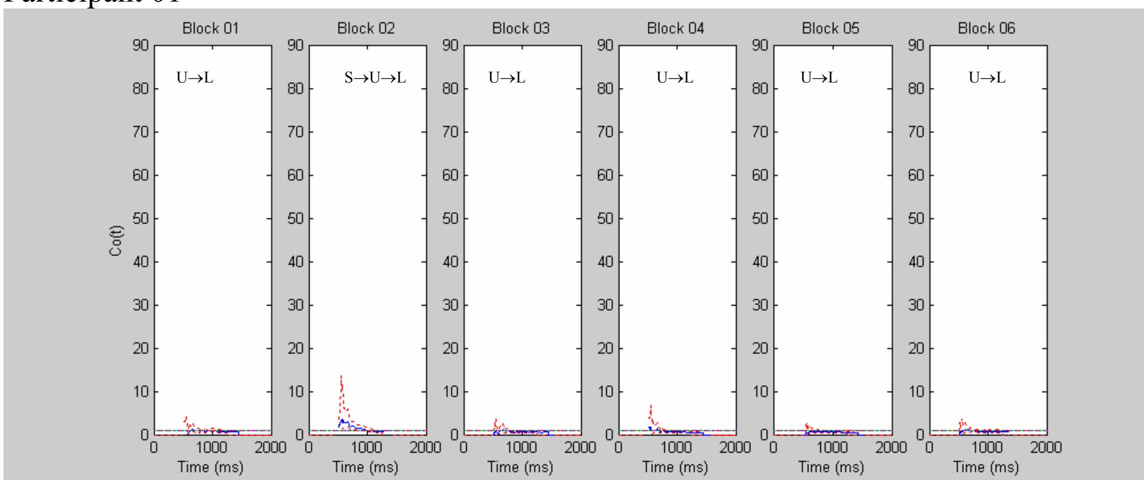
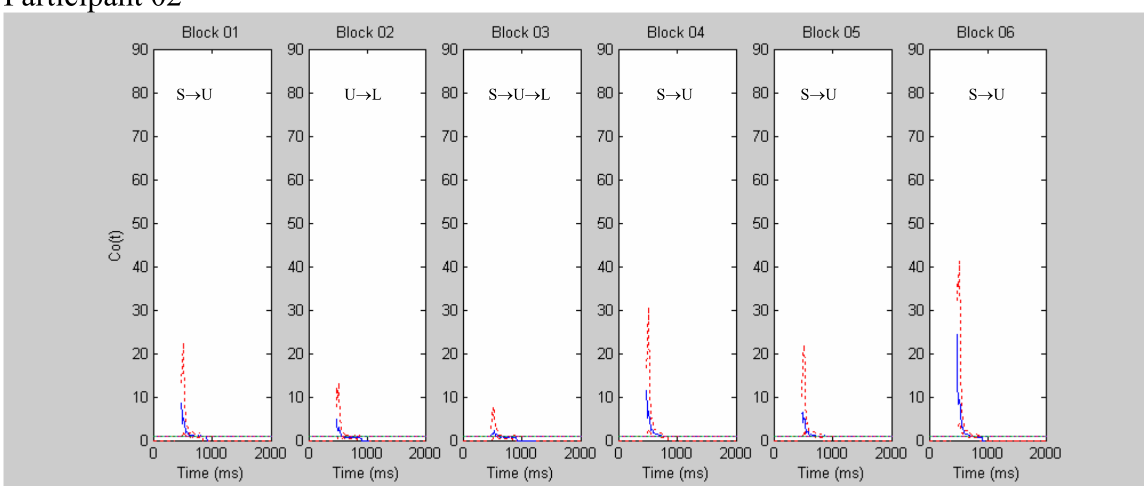


Figure 23: The **Absolute learning whole-mixed CCFs**, for the **Hero**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

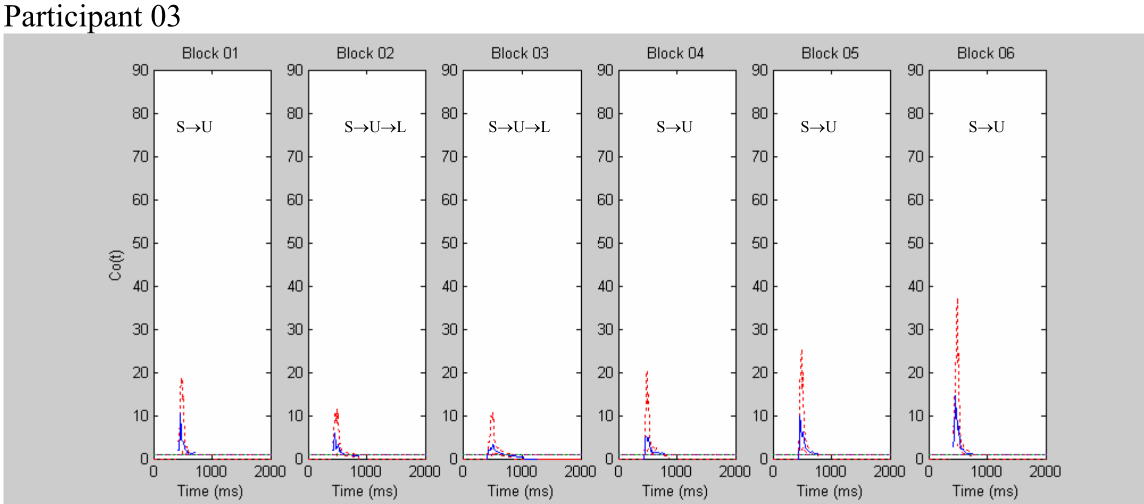
Participant 01



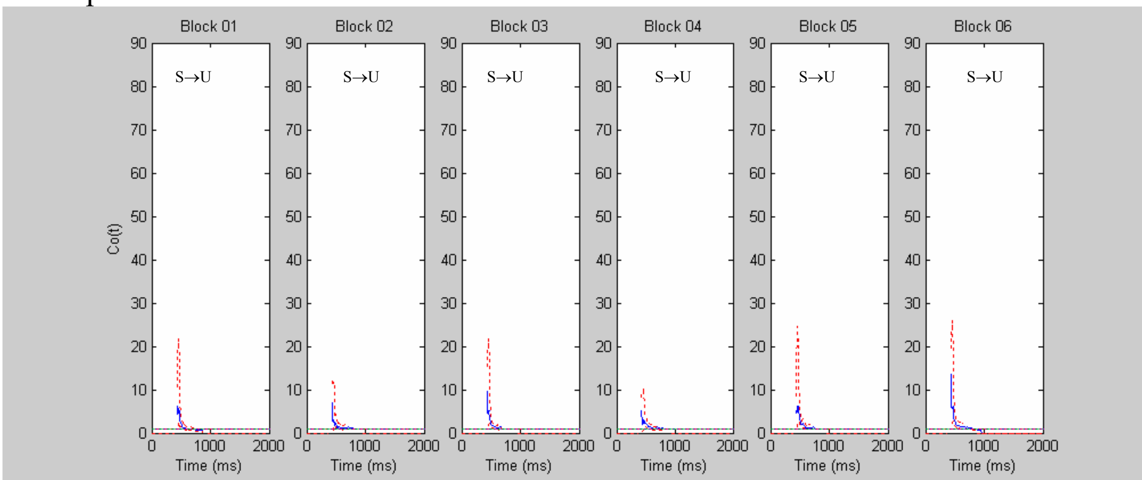
Participant 02



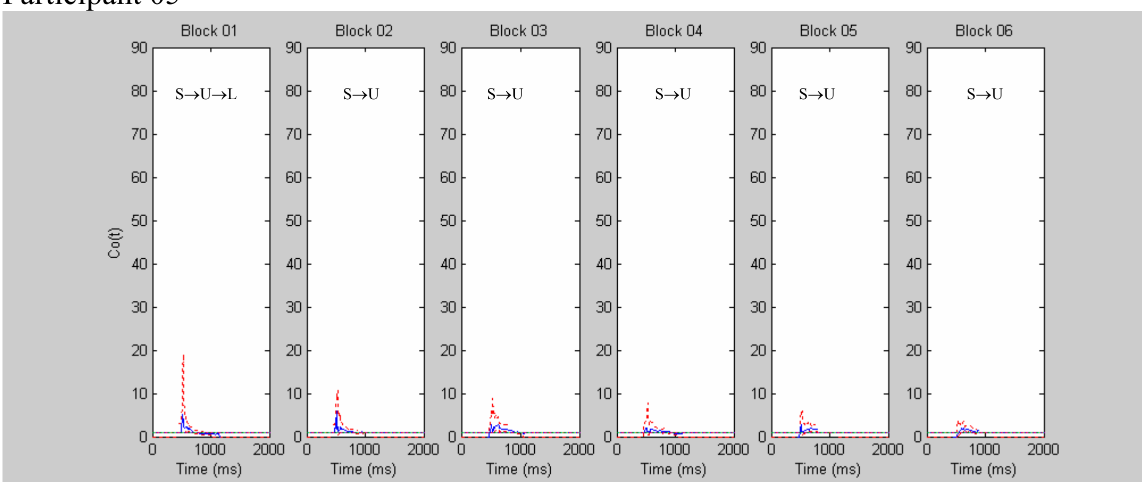
Participant 03



Participant 04



Participant 05



Participant 06

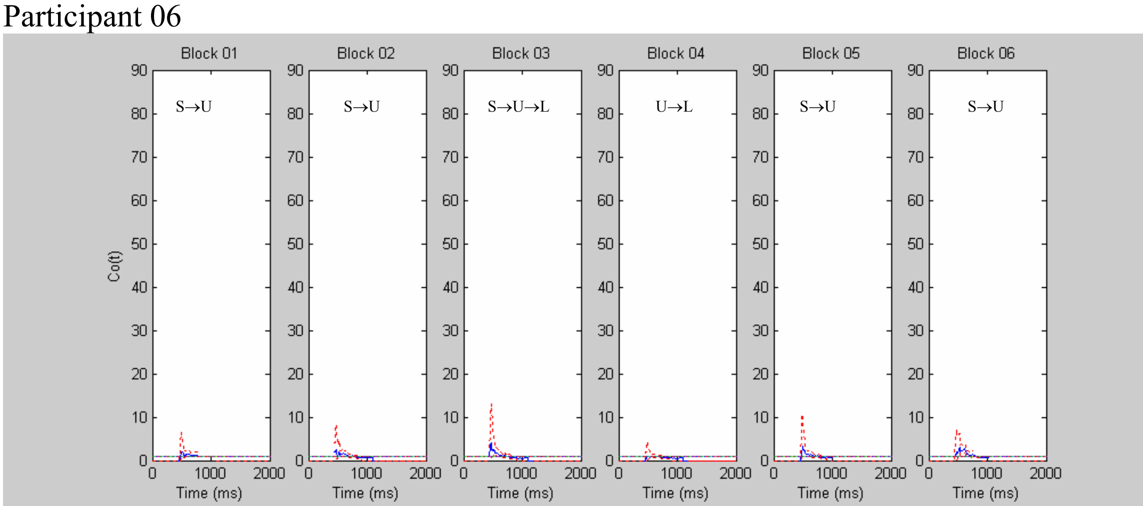
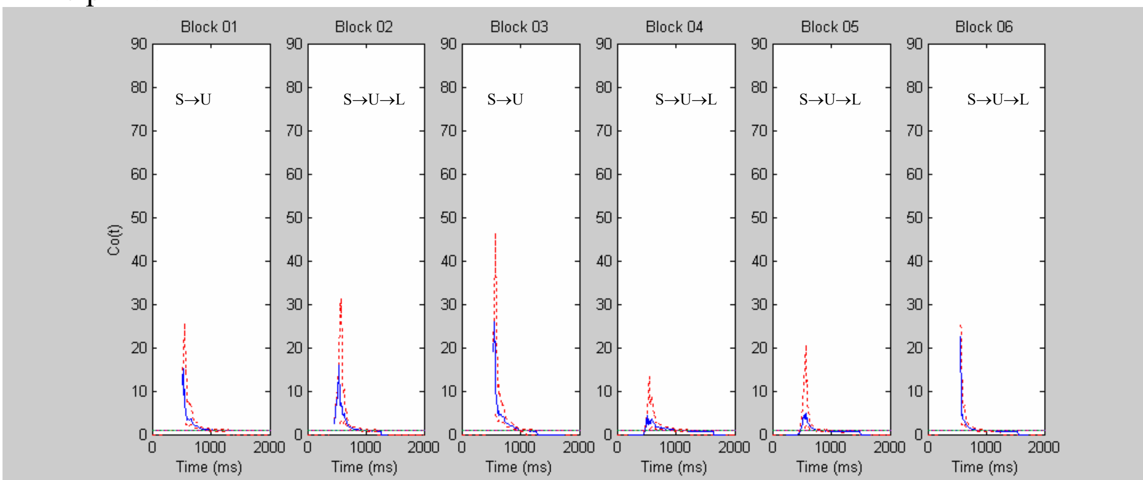
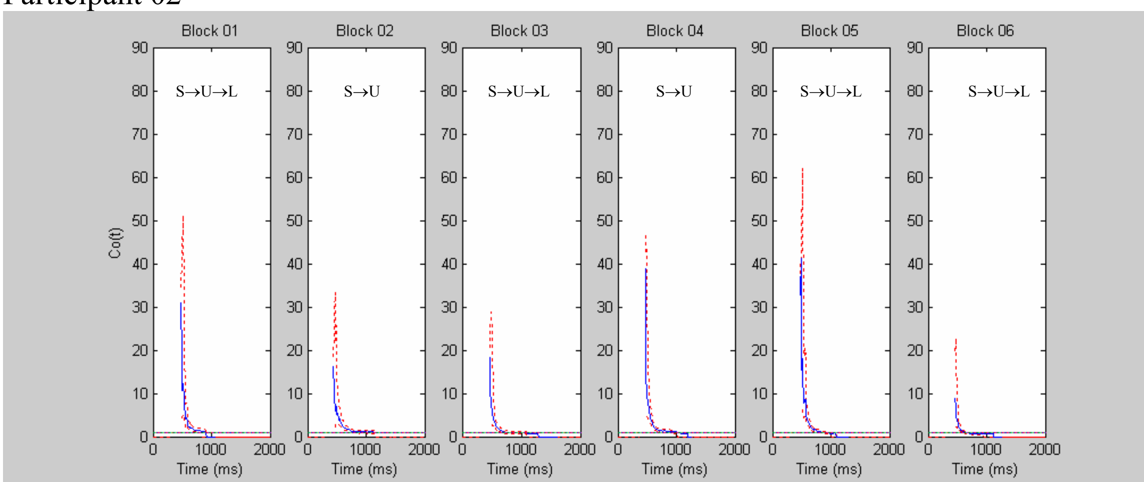


Figure 24: The **Relative learning whole-blocked CCFs**, for the **Hero**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

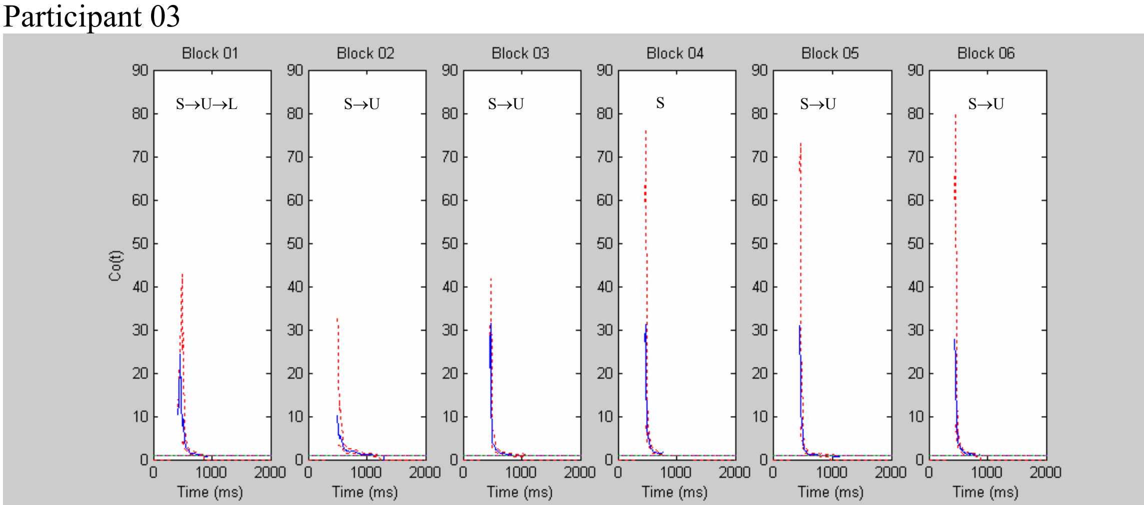
Participant 01



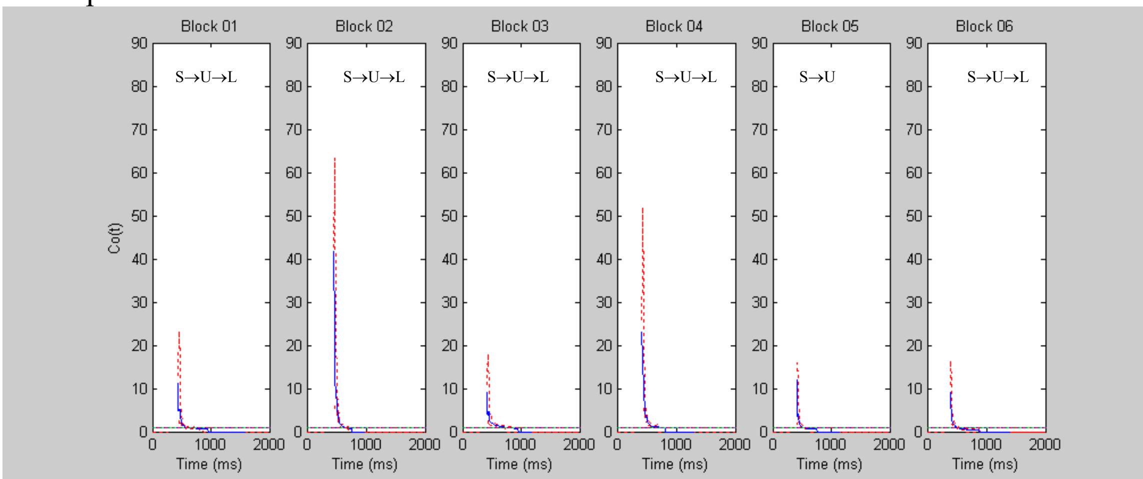
Participant 02



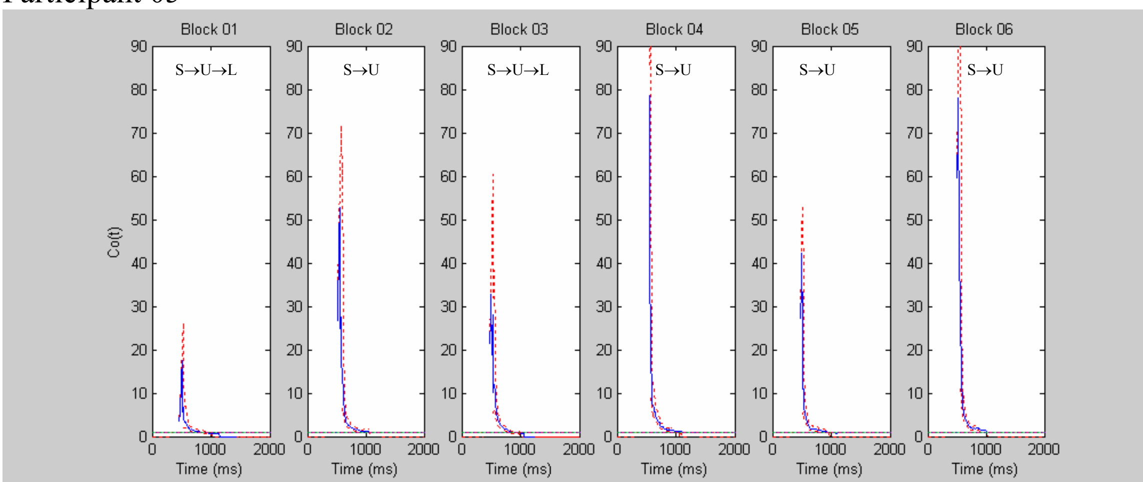
Participant 03



Participant 04



Participant 05



Participant 06

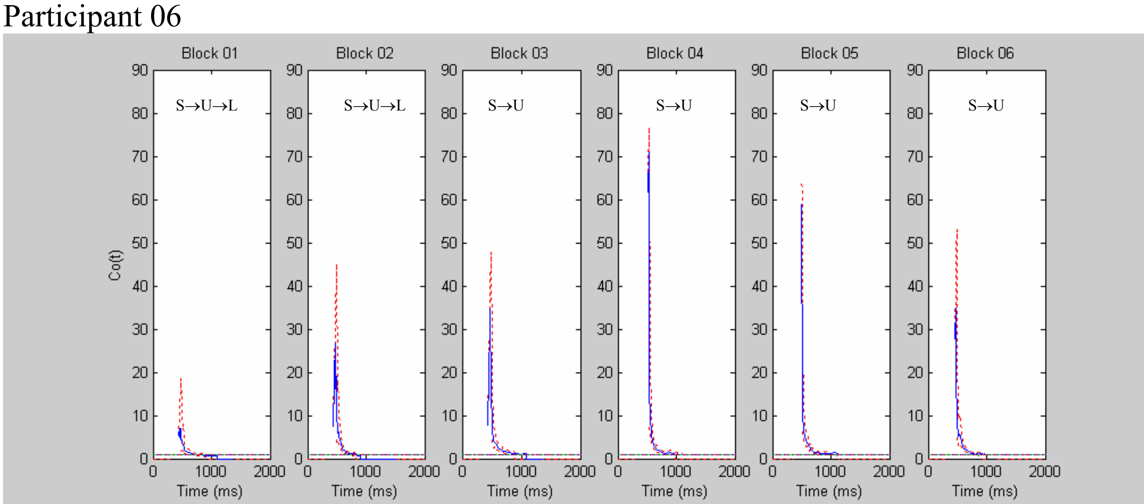
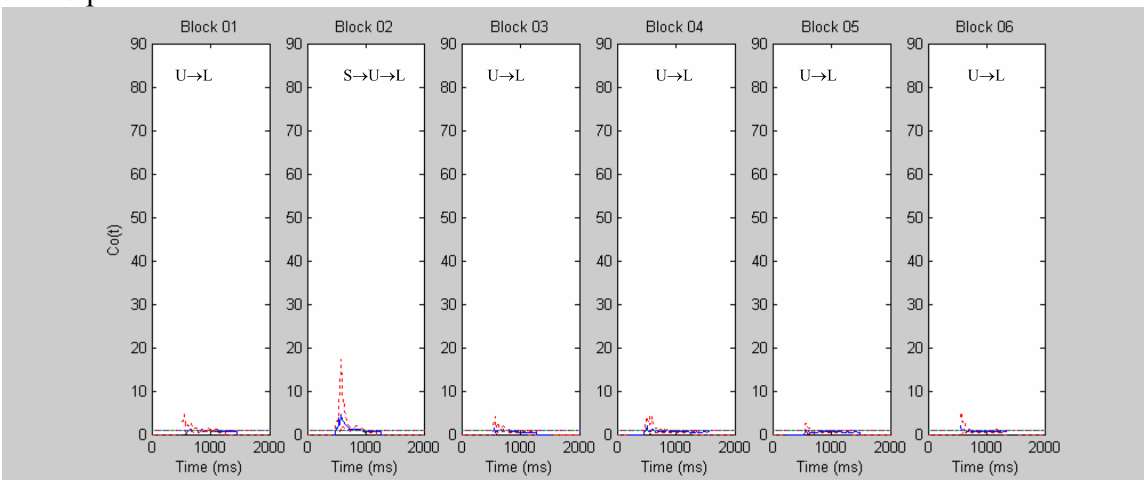
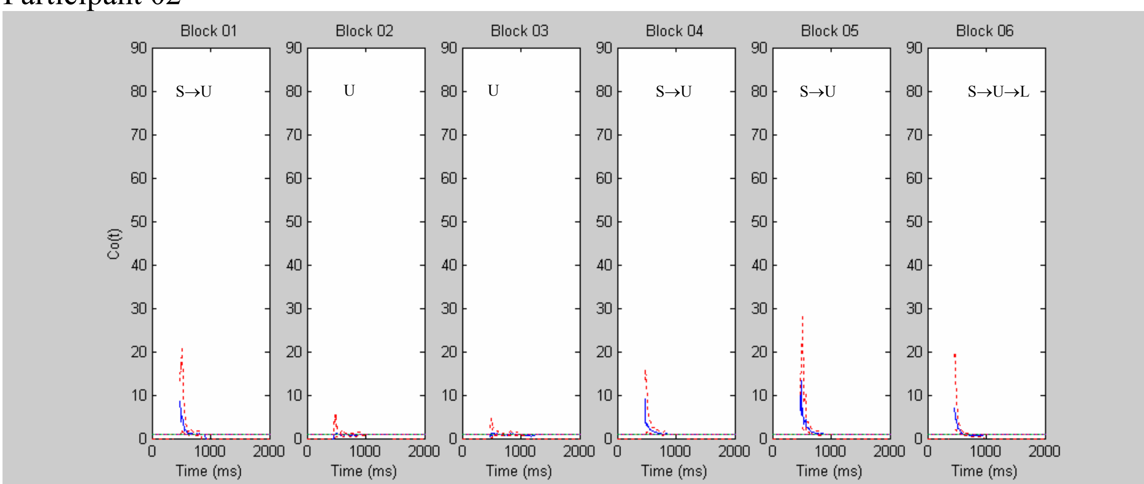


Figure 25: The **Relative learning whole-mixed CCFs**, for the **Hero**, across the learning sessions (Block), plotted separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity, then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

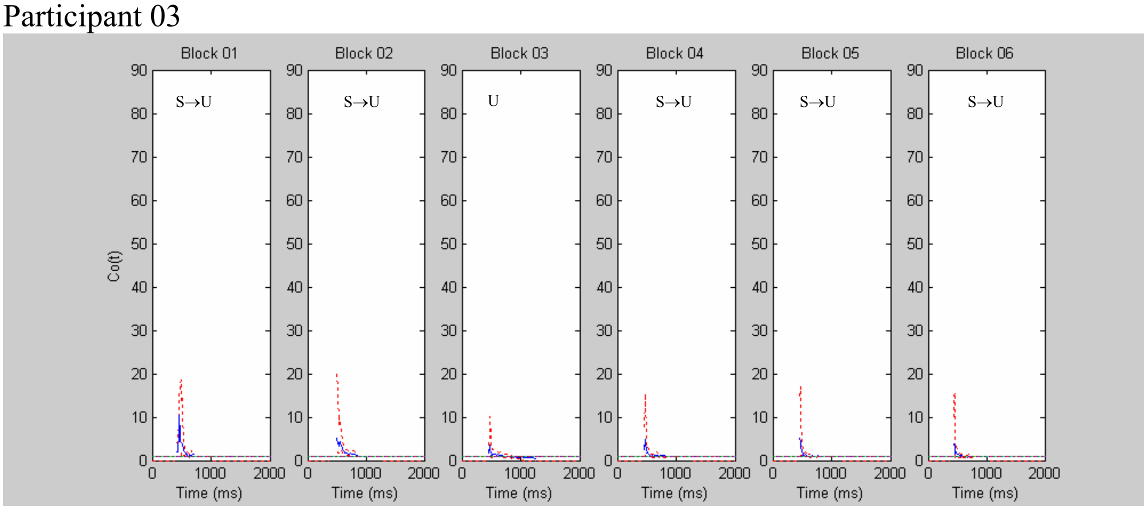
Participant 01



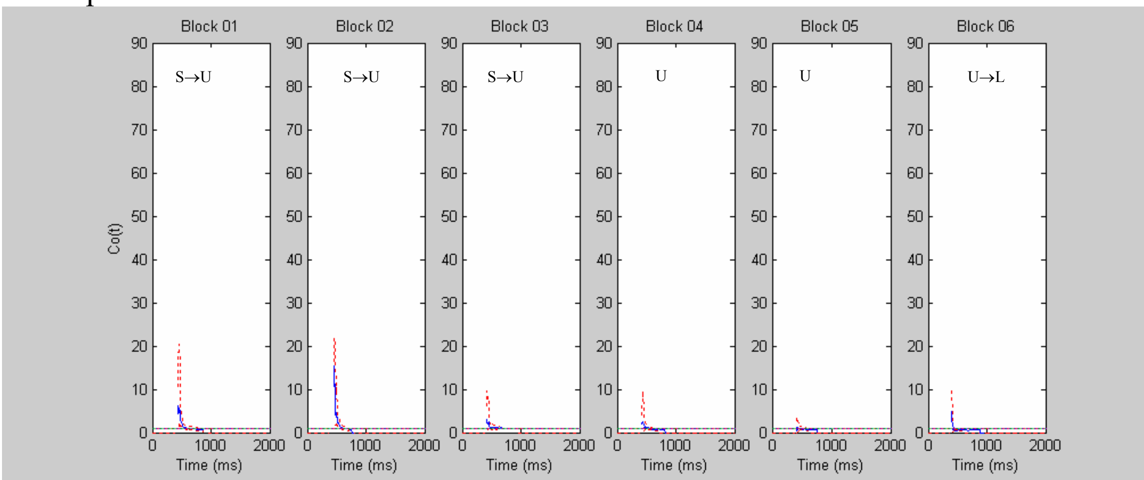
Participant 02



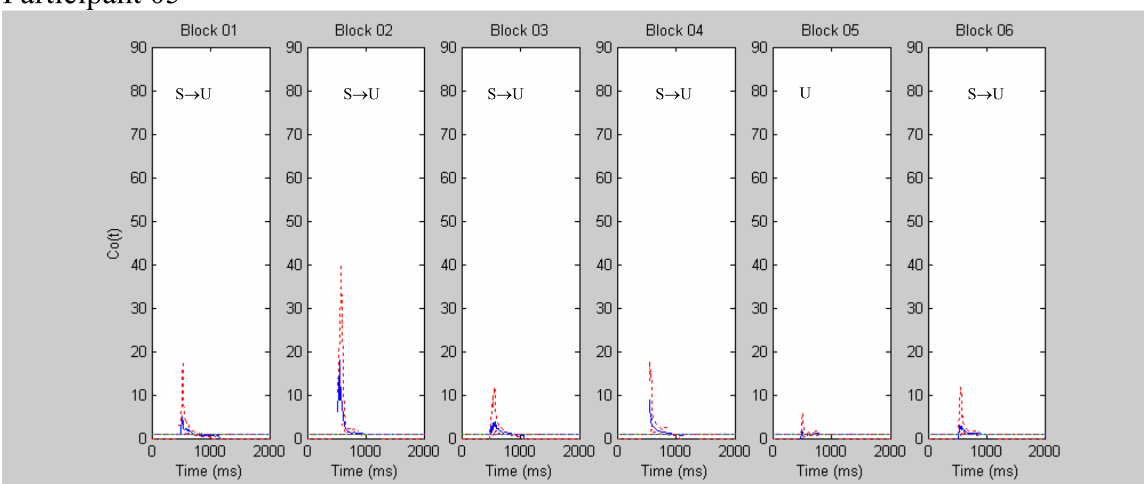
Participant 03



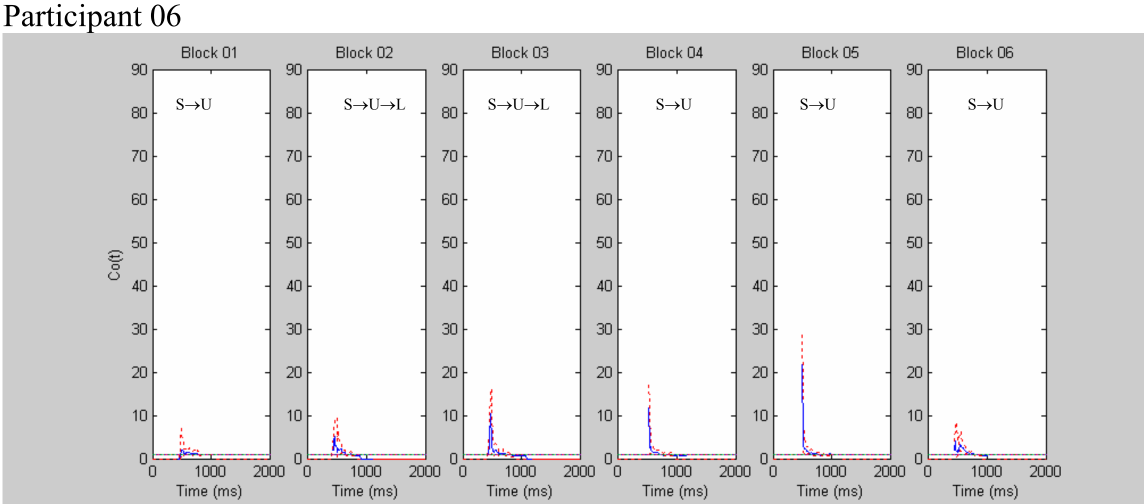
Participant 04



Participant 05



Participant 06



Analyses of Gang-members (which allows SFT test)

Overall, most of the CCFs violate C(t)=1 toward super capacity at early response times. The mixed block condition produced lower magnitude CCFs than the whole face blocked condition. This is consistent with the finding on mean RTs which showed slower processing in the mixed condition. Also the mixed condition exhibited unlimited CCFs in many cases. Both the absolute and relative learning CCFs clearly violate the capacity value of 1 toward super capacity in the blocked condition, for almost all participants and learning sessions. Also for some participants, there is a trend of magnitude change of CCFs over learning.

1. Absolute learning whole-blocked CCF (Figure 18): All participants, except 1 and 2 show increasing trend of CCF magnitude as a function of learning session. This increase reflects the amount of learning of the whole faces. Participants 1 and 2 exhibited a significant violation toward super capacity, with a flat learning trend.

2. Absolute learning whole-mixed CCF (Figure 19): Participants exhibited mainly unlimited CCF with a tendency to reach super capacity in later learning sessions (Participants 3, 4, 5, 6). Also, these participants showed an increase in magnitude of CCFs with learning. Participant 1 was super capacity for all sessions. Overall magnitudes of these CCFs were lower than the ones observed in the blocked condition. The resulting CCFs resembled those in (1) with smaller magnitude.
3. Relative learning whole-blocked CCF(Figure 20): In this condition all participants exhibited super capacity for all learning sessions. No monotonic trend of change in magnitude as a function of sessions is directly evident, across all participants.

On the whole, Participants 3, 5 and 6 exhibited an increasing trend to some extent. We can conclude that, for some participants, learning occurred at different levels for different sessions. For Participants 1, 2 and 4, it seems that the learning effects were largest at the beginning and in the middle of the learning session, followed by some decrease in learning activity.

4. Relative learning whole-mixed CCF (Figure 21): Super capacity has been demonstrated only in a few cases around sessions 5 and 6 for some participants. On the whole, unlimited capacity processing was demonstrated for almost all participants, with low magnitude of CCFs relative to the C(t)=1 bound. We can conclude that in the mixed condition participants were unable to benefit from the observation of whole faces only. No regular trend of CCF magnitude change can be observed across different participants, which is similar to the previous case (3).

Overall, super capacity dominated in the blocked condition, and most of the participants exhibited an increasing trend of absolute learning effects. The exhibited super capacity property suggests the presence of a possible positive interdependence between face features, which implicates the presence of gestalt properties during recognition of the gang-faces. Absolute and relative learning CCFs did not show the same trends over the learning sessions.

<u>Analyses of the capacity coefficient functions for the Hero face condition</u>

Again, note that the single hero face did not allow of application of the SFT tests for architecture and stopping rule. Like the gang faces, categorization in this case could be conducted using an OR rule, so we investigated capacity properties of the hero face as well.

Similar to the case of the gang faces, the overall finding is that most of the CCFs violate value C(t)=1 toward super capacity. The mixed condition produced a lower magnitude CCF than the blocked condition. In contrast to the gang faces the mixed condition exhibited super capacity CCFs, and the overall magnitude of the CCFs were larger. Both the absolute and relative learning CCFs clearly violate C(t)=1 toward super capacity in the blocked condition, for almost all participants and learning sessions. Also, for some participants there is a trend of CCF magnitude change.

1. Absolute learning whole-blocked CCF (Figure 22): All participants, except 1 and 2, show an increasing trend of CCF magnitude as a function of the earning session. That increase corresponds to the amount of learning of whole face. Participants 1 and 2 exhibited significant violation toward super capacity, with a flat learning trend.
2. Absolute learning whole-mixed CCF (Figure 23): Participants exhibited mainly super capacity CCFs. Also, Participants 2,3 and 4 showed increase in CCF magnitude of with learning, Overall magnitudes of CCFs were lower than the CCFs observed in the blocked condition (1).
3. Relative learning whole-blocked CCF (Figure 24): In this condition, all participants exhibited super capacity for all learning sessions. No monotonic trend of magnitude change as a function of sessions is directly evident across different participants. We can conclude that for some participants learning occurred at

different levels for different sessions. For Participants 1, 4 and 6 it seems the learning effects were strongest at the beginning and in the middle of learning sessions, followed by some decrease in learning activity.

4. Relative learning whole-mixed CCF (Figure 25): Super capacity has been demonstrated in many cases. We observed lower magnitude CCFs than in (3). No regular trend of CCF magnitude change can be observed across different participants, which is similar to case (3).

Overall, participants exhibited super capacity during perception of the hero face, and most of the participants exhibited an increasing trend of absolute learning effects. The super capacity property suggests the presence of possible positive interdependence between face features, which implies the presence of gestalt properties during recognition of the hero-faces.

Experiment test phase OR condition

Method

Participants

Same participants from the learning phase.

Materials

This experiment phase was divided into three different subexperiments: (1) Standard-face test session, identical to the face categorization session from the learning phase (2) Configural-face test and (3) Feature-face test (Figure 26). All participants performed each subexperiment 4 times, and the order of work for each participant was counterbalanced using the Latin-square design.

We will not provide here a detail description of the standard-face test, given that it was described in the previous section. The configural-face and feature-face tests had the same task design, except that they utilized different face stimuli. Thus, we will proceed with a generic description for both.

The instruction of both testing conditions resembled the standard-face test condition (see Figure 9). The only difference was that in both the first and second parts of the categorization task, new faces appeared in the set of stimuli, which were also tested with the SFT tests (the MIC and SIC) and capacity. Each subexperiment started with 24 practice trials.

Figure 26: The experiment test phase. This phase part was divided into three different subexperiments: (1) Standard-test session, identical to the one from the learning part; (2) Configural-test and (3) Featural-test. They all have three parts but they utilize different types of faces. In the second part, we used faces from the first part with the addition corresponding masked faces, for all subexperiments.

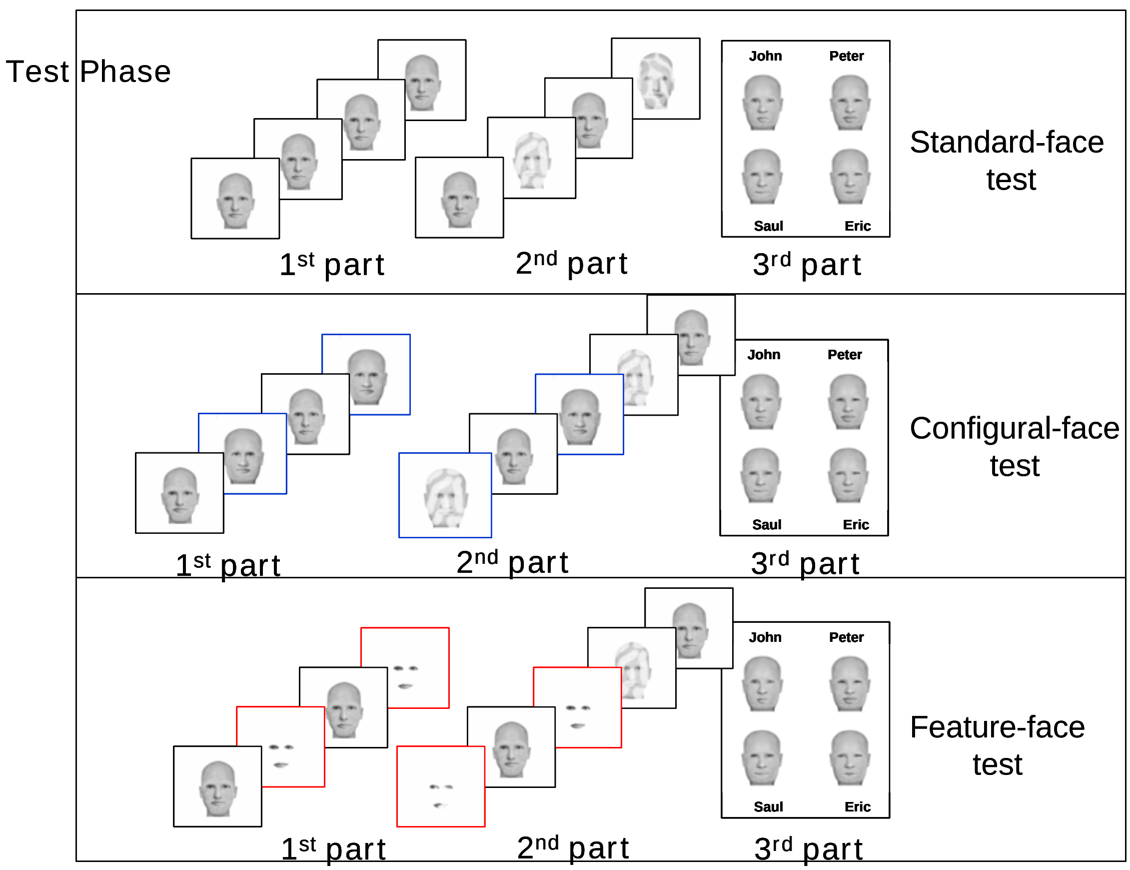
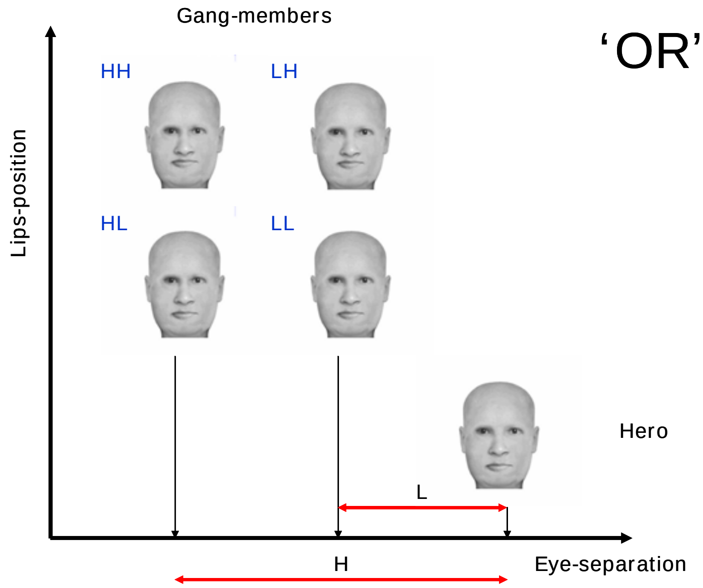
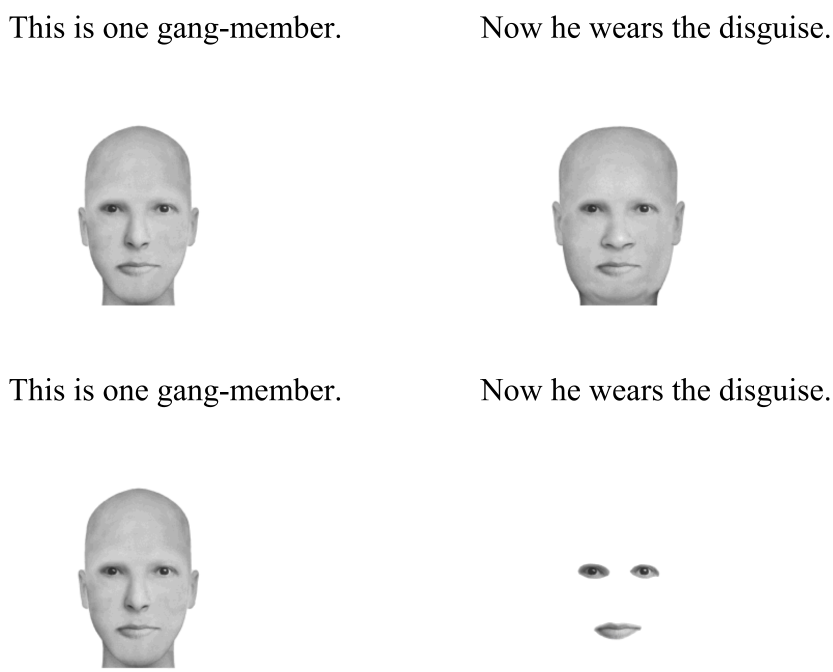


Figure 28: T**he OR condition, new faces in the configural-test**: two dimensional face-space defined by the eye-separation and lips-position. The saliency of features is defined by the marginal proximity of each face-feature projection onto the axes with respect to the other group members. In this case the projection, or marginal value, of each face feature for each gang-member is either close or distant with respect to the projection of the same feature from the hero. The design is the same as in the standard-test or the learning session except that the old faces have their face background completely changed.



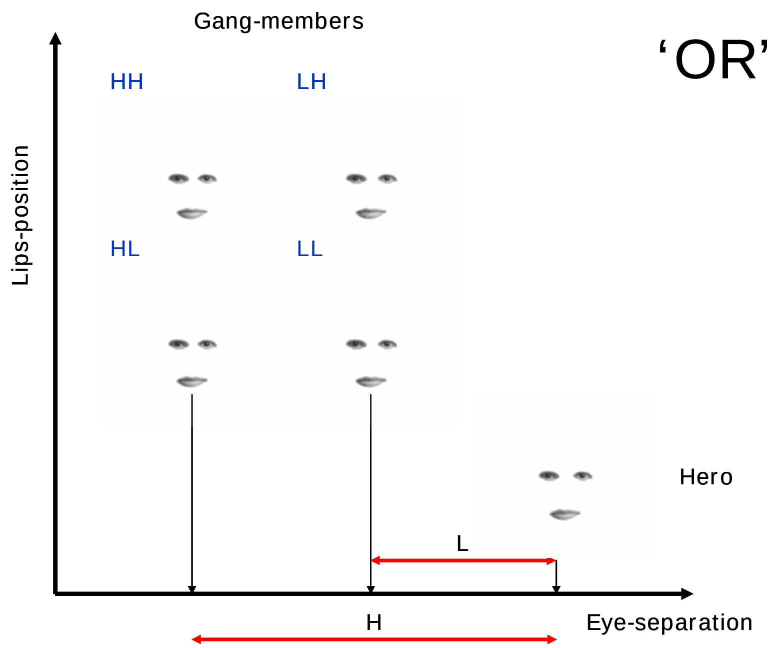
In the configural-test experiment (2) (Figure 28) participants were told the following story: "After an incident that happened with the gang and hero, the members from both groups are hiding from the police. We are informed that they put some disguises on their faces in order not to be recognized. However ALL of them wear the same disguise. The disguise covers everything except the eyes and lips. Have in mind that lips-position and eye-separation are the same as before because the disguise does not cover them" (Figure 27, 1st row).

Figure 27: Upper row: An old face and new face after configural manipulation. Note that the eye-separation and lips-position are the same, while the new face has the whole background of the face changed. Bottom row: In the feature faces (right) the face background has been removed.



In the featural-test experiment (3) (Figure 29) participants were told the following story: “After the incident they made the members from both the gang and hero are hiding from the police. We are informed that they put some disguises on their faces in order not to be recognized. The disguise covers everything except the eyes and lips. In this session some of the faces presented will have the eyes and lips only! Have in mind that the lips-position and eye-separation are the same as before because the disguise does not cover them” (Figure 27, 2^{nd} row).

Figure 29: **the OR condition, feature-faces in the featural-test**: two dimensional face-space defined by the eye-separation and lips-position. The saliency of features is defined by the marginal proximity of each face-feature projection onto the axes with respect to the other group members. In this case the projection, or marginal value, of each face feature for each gang-member is either close or distant with respect to the projection of the same feature from the hero. The design is the same as in the standard-test or the learning session except that the old faces have their face background completely removed.



So, in both subexperiments, the featural- and configural-test subexperiments, it was emphasized that the critical configural information, that was both necessary and sufficient to generate a correct response, did not change. In this part, participants received 400 trials of the old whole-faces, and 200 of the configurally manipulated faces. So for the SFT tests (MIC and SIC) there were 50 and 25 trials per condition, respectively in one session. Note that only gang faces are used in these analyses, which constitutes half of the total number of trials. The reason for using more old (standard) whole-face trials than new (configural whole-faces) stemmed from previous pilot studies which indicated that approximately 1:2 ratio can produce the biggest configural effects. We also wanted our participants to rely more on old whole-faces then to learn new faces to a larger extent. If they are exposed to new faces more than the old faces, they could learn them as new face stimuli, new face configurations, and then we will not be able to investigate the effect of disrupting of the old configurations.

After that portion in all subexperiments, participants received a set of trials designed for application of the capacity test (Figure 26, the 2nd part). Now, in subexperiments (2) and (3) both types of faces presented in the previous block were partially masked in order to examine the capacity of the system. So, all faces in both the old and new groups appeared 240 times for each group. Note that in this part 1/6 of the trials were old whole faces, 2/6 of the trials were old masked whole-faces, 1/6 trials were new configural whole-faces and the final 2/6 were new configural masked faces.

And, in the final part of all subexperiments participants finished with the with the four faces complete identification task with total of 100 trials (Figure 26, the 3rd part)

Design and procedure

Gang members were presented on half of the trials, while and the other half was the hero face. A participant had to decide, by pressing one of the mouse keys with the left and right index fingers, whether the hero or a gang-member was presented. RT was recorded from the onset of stimulus display, up to the time of response. Each trial consisted of a central fixation point (crosshair) for 1070ms followed by a high-pitch warning tone which lasted for 700msec. Then a face was presented for 190ms. Upon incorrect categorization, participants received an error message.

In the first part, participants received a total of 600 trials. The hero face was presented 300 times, as were all gang members, with each gang-member face was presented on 75 trials. In the second part, participants observed another 240 trials, of which 80 belonged to both types whole faces (divided equally between the hero and gang faces), and the other 160 trials belonged to the same two groups but with different types of masks (80 trials for each mask) applied on both groups.

In the third part, the experiment a complete identification task where participants had to learn to identify 4 faces, a total of 100 trials were used. We used same presentation rates as in the previous parts.

The trial presentation order was pseudo-randomized within each session. In each session, participants ran 5 blocks of approximately 200 trials each. All participants accomplished a total of 12 sessions (three parts for sessions each) during the test phase of the experiment. The participants were instructed to achieve high accuracy and to respond as fast as possible. For the analyses we aggregated the four experimental sessions into one, therefore achieving larger number of observation per condition. The reason for the data aggregation was to attain more statistical power. So for the SFT tests, each factorial condition (HH, HL, LH and LL) possessed around 600 trials. And for the capacity analysis, we provided approximately 320 trials for each integrated hazard function, in each subexperiment.

Results

Basic Mean RT Analyses

First, we compared mean processing time for the different subexperiments: the configural-test, featural-test and standard-test, for all the gang and hero trials together, in the experiment. We also performed analyses for the gang-members on blocked trials, separately. All participants exhibited the following ordering: $RT_{standard}<RT_{feature}<RT_{configural}$, except Participant 1 (Table 3 left panel – "All trials"). The main effect of the type of experiment was significant for all separate participants, at the level of $p<0.01$, with power=1. So, almost all participants were fastest in the standard-test conditions with the old faces, than in both the new configural-test and featural-test subexperiments. Further, the experiment with the new configuration faces exhibited slower processing than the experiment when feature faces were used. It could be suggested that on average, the new face context presented in the configural subexperiment produced the most detrimental effect on face perception.

We also investigated the mean difference on processing of old whole faces in the different subexperiments. In the configural experiment, old faces were mixed with the new configural faces; in the featural experiment they were mixed with the faces that contained only features, and in the third experiment, or the standard test, they were presented alone, as in the learning phase. As in the previous analysis on all trials, almost all participants exhibited the RT ordering $RT_{standard}<RT_{feature}<RT_{configural}$, except Participant 1 who exhibited $RT_{standard}<RT_{configural}<RT_{feature}$, ($F(1,1948)= 14.847$, $p<0.01$); and Participant 4 who exhibited a significant difference, with $RT_{standard}>RT_{feature}$, ($F(1, 1976)=5.61$, $p<0.05$).

Table 3: Mean RTs and standard errors for each subexperimental condition (configural-test, featural-test and standard-test), for individual participants. Table is vertically divided into two parts: all experimental trials were averaged for both hero and gang faces on the left side; and averaged over gang-member blocked trials, which are old whole faces from different subexperiments, on the right side.

	All Trials		OLD Gang-members blocked trials	
	Mean	Std. Error	Mean	Std. Error
Participant 01	RT(ms)		RT(ms)	
Configural-test	743.503	3.710	701.830	7.888
Featural-test	781.695	3.730	739.109	7.816
Standard-test	791.344	4.275	742.707	6.407
Participant 02				
Configural-test	595.072	2.641	564.991	5.848
Featural-test	590.354	2.669	530.095	5.779
Standard-test	551.063	3.043	510.442	4.710
Participant 03				
Configural-test	553.857	1.912	507.453	3.067
Featural-test	503.270	1.900	457.305	3.059
Standard-test	480.206	2.197	444.177	2.502
Participant 04				
Configural-test	507.224	1.847	493.541	3.590
Featural-test	486.697	2.057	452.357	3.975
Standard-test	477.183	2.373	463.747	3.249
Participant 05				
Configural-test	670.019	2.715	595.790	4.707
Featural-test	629.727	2.717	549.622	4.700
Standard-test	578.045	3.097	505.953	3.830
Participant 06				
Configural-test	560.216	2.339	515.668	4.266
Featural-test	560.624	2.358	490.759	4.190
Standard-test	528.277	2.669	484.826	3.424

From this analysis, we can conclude that mean processing time of the old whole-faces is different for different experimental contexts, and is differently affected by the presence of either configural or featural faces. We will use this finding later, when calculating the capacity coefficients for different conditions.

In the subexperiments 1 and 2 we used new changed faces and old whole-faces (Figures 28 and 29): in experiment 1 (configural test) we replaced the complete face background except for the two important face dimensions (eyes and lips). In the experiment 2 (featural test) we completely removed the face background and left the only two critical face properties. In both experiments, we are interested in whether these manipulations produced changes in the processing organization, i.e. the architecture of mental processes during face recognition. So the outcome of each face manipulation was compared to the outcome of the processing of the old faces. But the old faces were presented in multiple experimental situations: in the standard-test experiment, which was the exact copy of the design used in the learning phase, as well as in each subexperiment, together with the configurally altered faces (configural- and featural-test). Since we demonstrated that mixing different types of faces could produce changes in processing of each type, (the difference between the standard-test experiment and each of the two: the configural- and featural-test) we investigated architectural differences between processing of each type of altered faces (configural and featural) and the processing of old faces in all that experiments.

Comparison of the processing characteristics between old faces and new faces for the standard- and configural-tests (Configural x Standard design)

We ran the GLM analysis, type I model, using the following fixed factors: the eye-separation (high/low), lips-position (high/low) and the experimental group (3 levels: group1=old-configural, group2=new-configural and group3=old-standard). The old-configural conditions are based on trials when old faces from the configural-test were used. The new configural conditions consisted of new faces made from the old faces by configural alteration (changing the face context), also from the configural-test. The old standard faces are taken from the standard-test. As a covariate we chose the trial order. Of the most interest in this study is the relation between groups 1 and 2 and between groups 1 and 3.

In both groups the old-configural and old-standard, we examined processing of whole faces, but the difference is that in the old-standard, whole faces were not mixed with configurally altered faces. Therefore we expect that the mean face processing should be fastest in the old-standard condition (group 1). We expect that the main effect of the experimental group will be significant, and that the three way interaction between two face-features of interest and experimental group (Eyes x Lips x Exp group) will be significant. The interaction would indicate a possible change in the architecture between different experimental groups. Namely, we expect that processing of the old-standard faces could be based on a different mechanism than processing of the new-configural faces. Given that, on average, we observed differences between the processing of whole faces when they are presented along and when combined with configurally altered faces (the old-standard and old-configural faces) we expect to observe some changes in the processing architecture. In fact, we expect that by altering the face configuration the system will switch from a fast, efficient processor (probably parallel) to a less optimal processing system (maybe serial) under the constraints of new face configuration.

Results from the GLM analyses are presented in Table 4 (left) and the mean RTs are presented in the Figure 30.

Figure 30 : Mean RTs from the test phase for different types of faces. The old and new configural faces are from the configural-test; the old and new featural faces are from the featural-test. The old faces are from the standard-test subexperiment. Error bars around the mean RT indicate standard error.

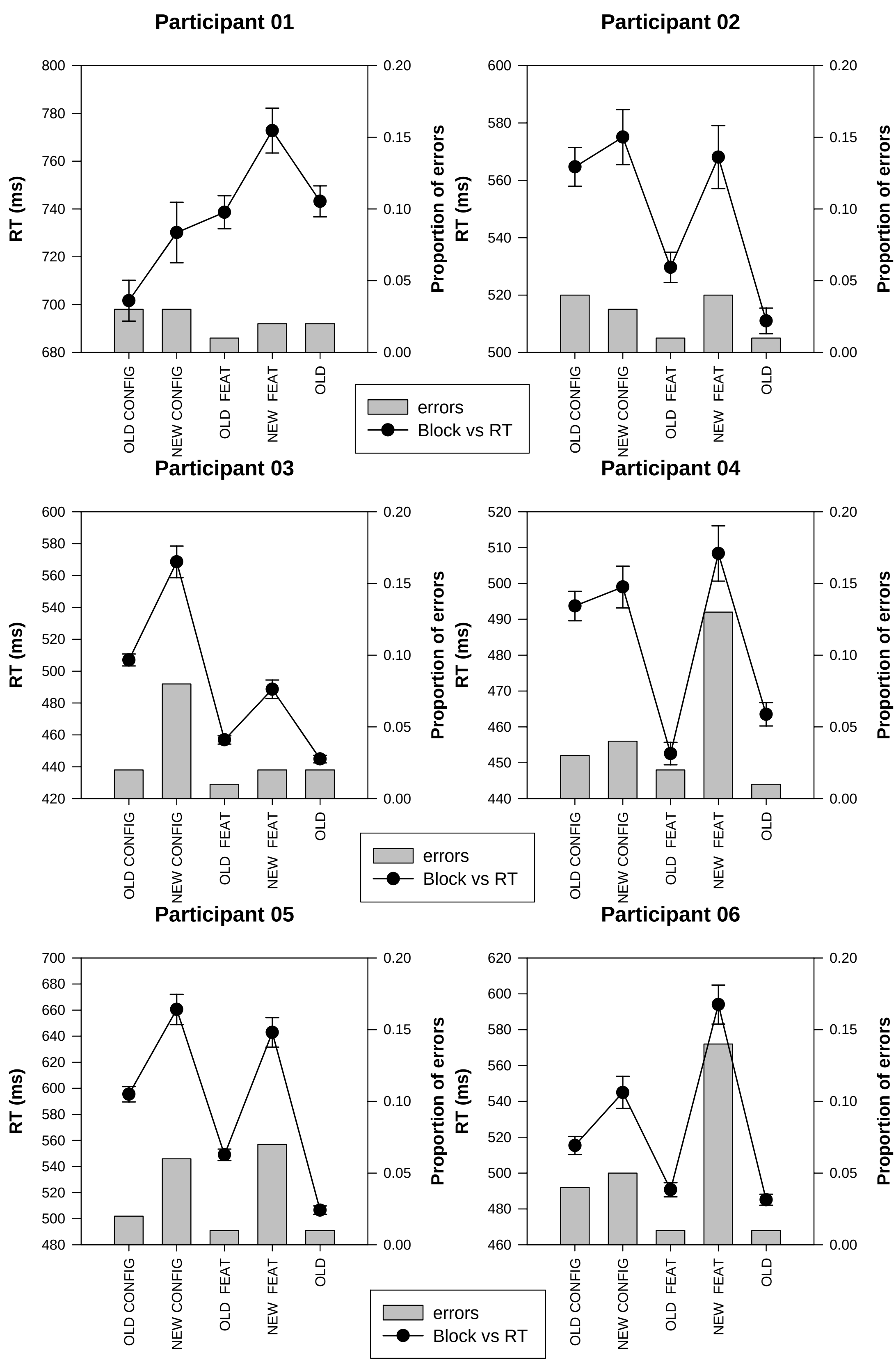

Participant 01
Participant 02
Participant 03
Participant 04
Participant 05
Participant 06
RT (ms)
Proportion of errors
OLD CONFIG
NEW CONFIG
OLD FEAT
NEW FEAT
OLD
errors
Block vs RT

Table 4: GLM univariate analysis was conducted on gang-members, for individual participants (SUB). The main effects and interaction terms are listed in the second column (*Factor*). Degrees of freedom are in the third column (*df*). The *error* row defines the degrees of freedom for the F-test error term, for that participant. Each F-test value (*F*) has two degrees of freedom: one from its corresponding row, and the second one from the error row. A significance level is presented in the column *Sig.*, and the observed power for that effect is in the last column. Table is vertically divided between two different analyses: a comparison of the processing characteristics between old faces and new faces for the standard- and configural-tests (Configural x Standard design, on the left side) and a comparison of the processing characteristic differences between old faces and new faces based on removing the face context (Featural x Standard design, on the right side)

Configural x Standard	df	F	Sig.	Observed Power	**Featural x Standard**	df	F	Sig.	Observed Power
Participant 01									
Trial order	1	6.711	.010	.736	**Trial order**	1	4.147	.042	.530
Eyes	1	176.247	.000	1.000	**Eyes**	1	160.531	.000	1.000
Lips	1	124.336	.000	1.000	**Lips**	1	113.848	.000	1.000
Exp group	2	8.001	.000	.956	**Exp group**	2	4.723	.009	.791
Eyes x Lips	1	78.816	.000	1.000	**Eyes x Lips**	1	50.201	.000	1.000
Eyes x Exp group	2	3.685	.025	.679	**Eyes x Exp group**	2	1.792	.167	.376
Lips x Exp group	2	1.823	.162	.382	**Lips x Exp group**	2	7.514	.001	.944
Eyes x Lips x Exp group	2	1.023	.360	.230	**Eyes x Lips x Exp group**	2	3.198	.041	.613
Error	2326				**Error**	2344			
Participant 02									
Trial order	1	11.224	.001	.918	**Trial order**	1	.965	.326	.166
Eyes	1	278.019	.000	1.000	**Eyes**	1	231.284	.000	1.000
Lips	1	139.735	.000	1.000	**Lips**	1	72.244	.000	1.000
Exp group	2	44.900	.000	1.000	**Exp group**	2	21.946	.000	1.000
Eyes x Lips	1	79.190	.000	1.000	**Eyes x Lips**	1	47.898	.000	1.000
Eyes x Exp group	2	4.170	.016	.736	**Eyes x Exp group**	2	7.345	.001	.939
Lips x Exp group	2	.378	.685	.111	**Lips x Exp group**	2	14.954	.000	.999
Eyes x Lips x Exp group	2	3.384	.034	.639	**Eyes x Lips x Exp group**	2	6.145	.002	.891
Error	2336				**Error**	2348			
Participant 03									
Trial order	1	50.089	.000	1.000	**Trial order**	1	27.422	.000	.999
Eyes	1	450.579	.000	1.000	**Eyes**	1	470.282	.000	1.000

Lips	1	383.481	.000	1.000	**Lips**	1	174.005	.000	1.000
Exp group	2	338.709	.000	1.000	**Exp group**	2	55.060	.000	1.000
Eyes x Lips	1	242.839	.000	1.000	**Eyes x Lips**	1	102.546	.000	1.000
Eyes x Exp group	2	42.307	.000	1.000	**Eyes x Exp group**	2	22.960	.000	1.000
Lips x Exp group	2	55.523	.000	1.000	**Lips x Exp group**	2	4.442	.012	.764
Eyes x Lips x Exp group	2	35.546	.000	1.000	**Eyes x Lips x Exp group**	2	2.354	.095	.478
Error	2320				**Error**	2351			
Participant 04									
Trial order	1	10.040	.002	.886	**Trial order**	1	1.278	.258	.204
Eyes	1	283.332	.000	1.000	**Eyes**	1	168.151	.000	1.000
Lips	1	46.173	.000	1.000	**Lips**	1	44.254	.000	1.000
Exp group	2	27.342	.000	1.000	**Exp group**	2	39.638	.000	1.000
Eyes x Lips	1	21.027	.000	.996	**Eyes x Lips**	1	29.968	.000	1.000
Eyes x Exp group	2	12.406	.000	.996	**Eyes x Exp group**	2	11.672	.000	.994
Lips x Exp group	2	5.152	.006	.827	**Lips x Exp group**	2	.459	.632	.125
Eyes x Lips x Exp group	2	1.411	.244	.304	**Eyes x Lips x Exp group**	2	2.751	.064	.545
Error	2624				**Error**	2340			
Participant 05									
Trial order	1	22.886	.000	.998	**Trial order**	1	2.493	.114	.352
Eyes	1	199.115	.000	1.000	**Eyes**	1	320.334	.000	1.000
Lips	1	143.791	.000	1.000	**Lips**	1	38.270	.000	1.000
Exp group	2	229.018	.000	1.000	**Exp group**	2	181.158	.000	1.000
Eyes x Lips	1	47.818	.000	1.000	**Eyes x Lips**	1	32.382	.000	1.000
Eyes x Exp group	2	6.366	.002	.901	**Eyes x Exp group**	2	47.095	.000	1.000
Lips x Exp group	2	26.713	.000	1.000	**Lips x Exp group**	2	10.002	.000	.985
Eyes x Lips x Exp group	2	4.822	.008	.800	**Eyes x Lips x Exp group**	2	3.870	.021	.702
Error	2341				**Error**	2341			
Participant 06									
Trial order	1	3.881	.049	.504	**Trial order**	1	7.079	.008	.758
Eyes	1	212.241	.000	1.000	**Eyes**	1	183.303	.000	1.000
Lips	1	106.000	.000	1.000	**Lips**	1	109.716	.000	1.000
Exp group	2	44.545	.000	1.000	**Exp group**	2	146.558	.000	1.000
Eyes x Lips	1	56.135	.000	1.000	**Eyes x Lips**	1	66.053	.000	1.000
Eyes x Exp group	2	5.181	.006	.829	**Eyes x Exp group**	2	19.532	.000	1.000
Lips x Exp group	2	1.842	.159	.386	**Lips x Exp group**	2	3.693	.025	.680
Eyes x Lips x Exp group	2	.651	.522	.160	**Eyes x Lips x Exp group**	2	.333	.717	.104
Error	2327				**Error**	2314			

From the results presented in Table 4 (left part of the table), we see that all main effects of Lips and Eyes were significant at the $p<0.01$ level, with power=1. Also the main effect of experimental group was significant for all participants, which indicates that there is a significant mean RT difference between experimental groups. In Figure 30 we present mean RTs for different experimental groups: all participants except Participant 1, exhibited the following order of mean RTs: $RT_{old\text{-}faces}<RT_{old\text{-}configuration}<RT_{new\text{-}configuration}$ (old-faces are whole faces from the standard condition, and old-configuration faces are the old faces presented together with the configurally altered faces, and new configuration are the new-faces). Participant 1 exhibited reverse trend for the old-standard and new-configural faces, i.e. he exhibited slower mean RT processing on whole faces when they were not combined with the configurally altered faces. On the other hand, he demonstrated (as other participants did) faster processing of whole faces than configural faces when they were combined together in the experimental sessions.

Additionally, the MIC tests, revealed by the two-way interaction Lips x Eyes, were significant for all participants, with very high power>.99. Overall, MIC values were positive for all participants, revealing oveardditivity. However, further analysis of the MIC values, will be broken down for different experimental conditions, for each participant (see Figures 31, 32 and 33).

Figure 31: The SFT test results for the OR **standard-test** condition, for gang-member faces, for all participants. The results are based on all sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.

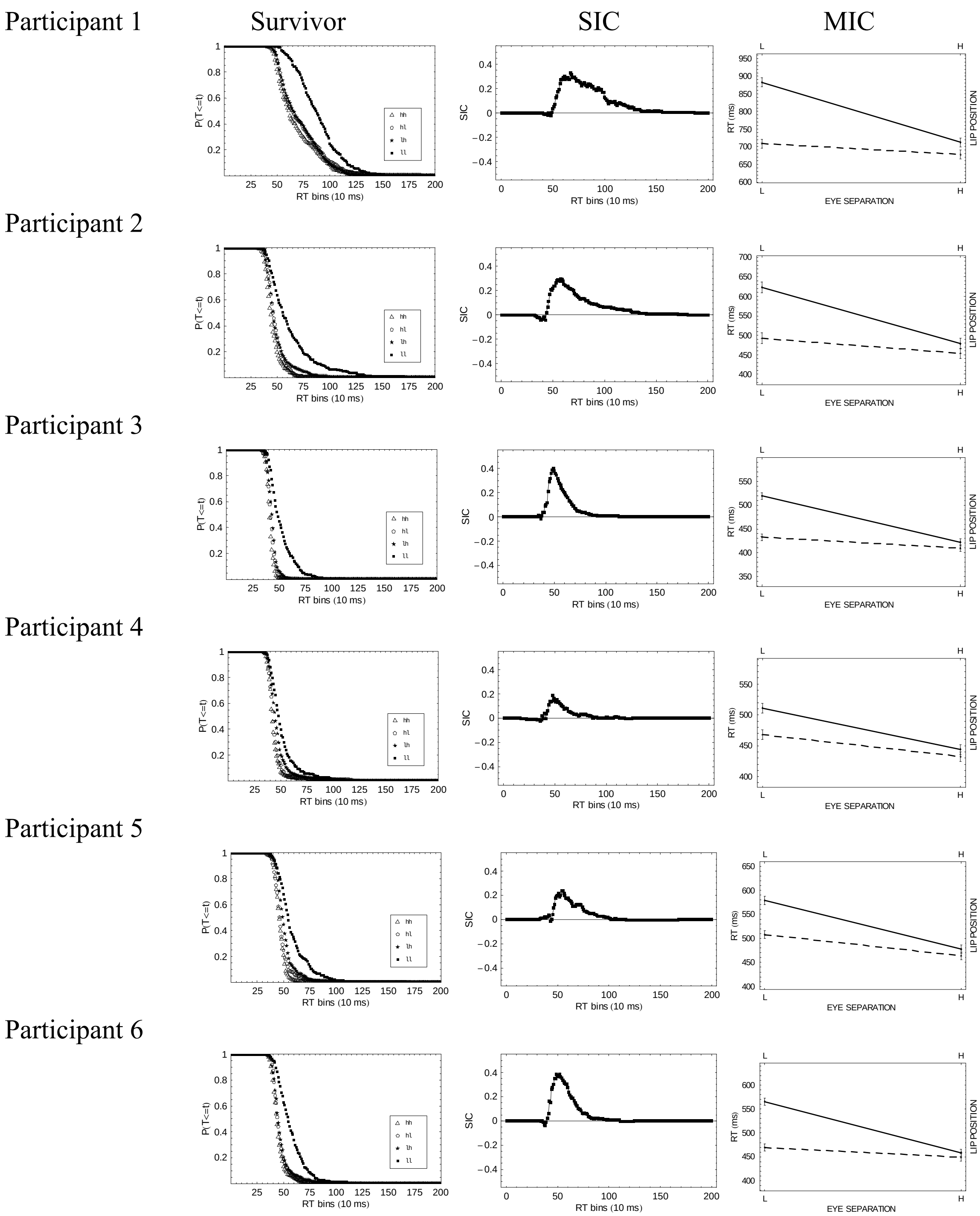


Figure 32: The SFT test results for the OR **configural-test** condition, for gang-member faces, for all participants. The results are based on all sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.

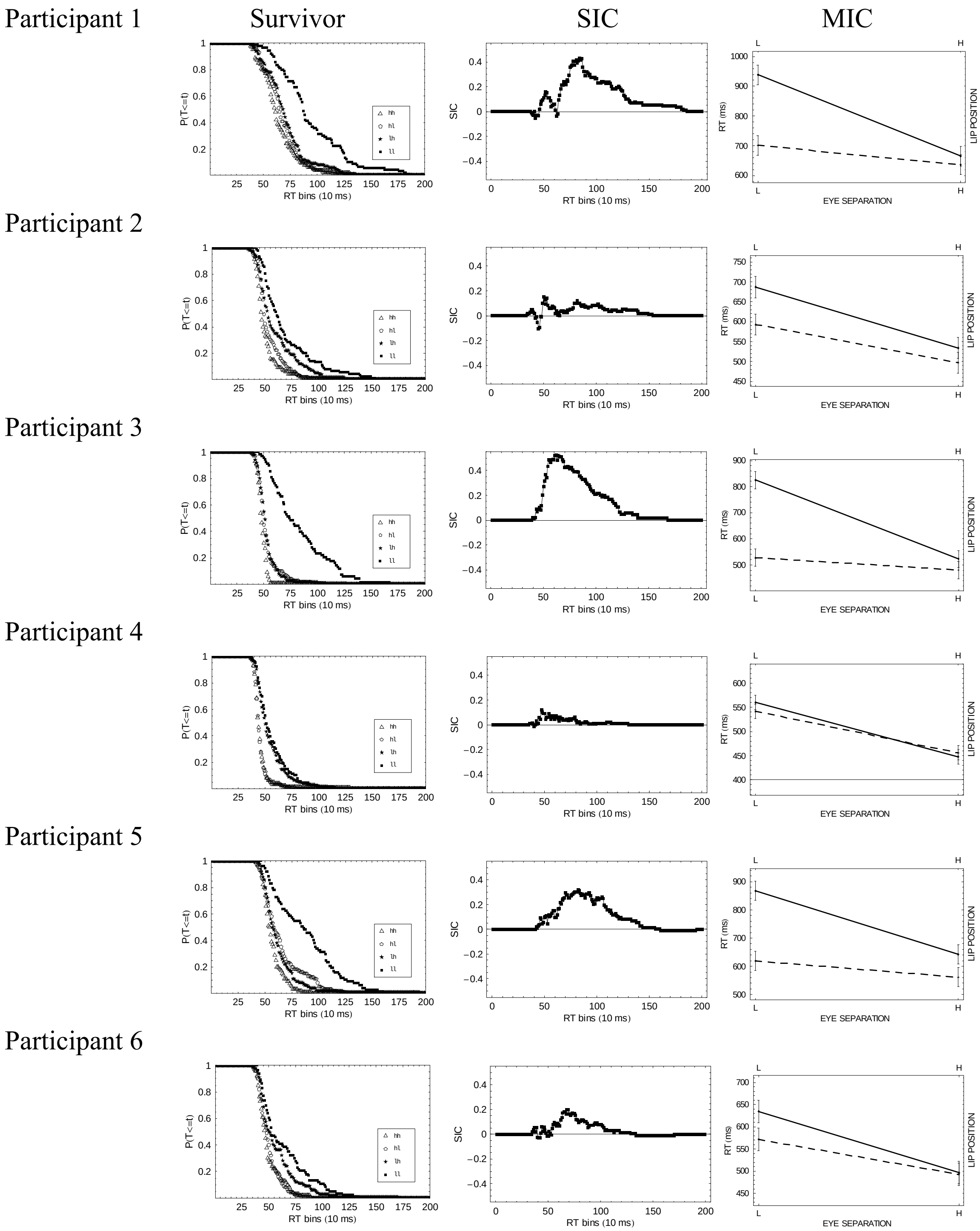
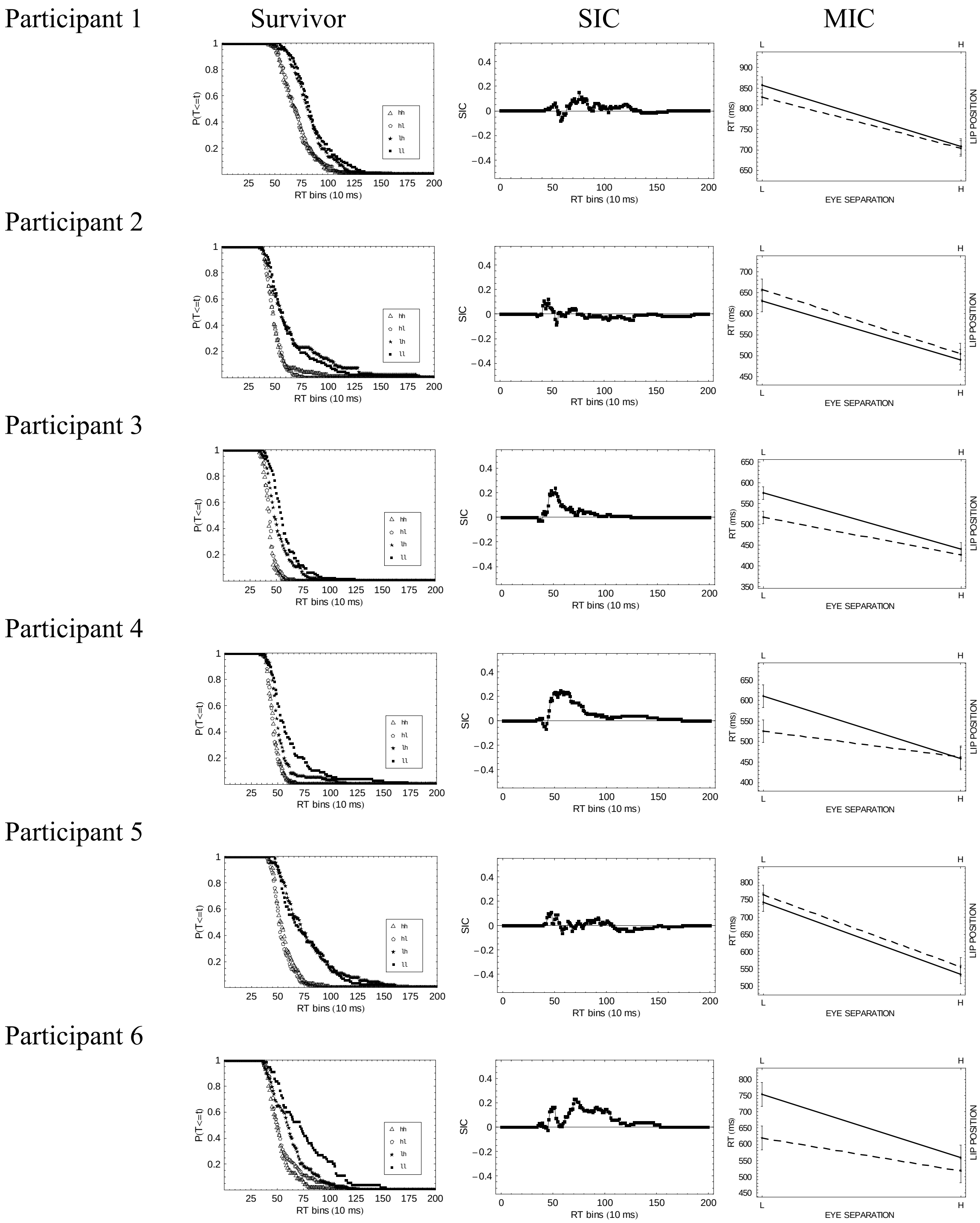


Figure 33: The SFT test results for the OR **featural-test** condition, for gang-member faces, for all participants. The results are based on all sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.



The three way interaction Eyes x Lips x Exp group, indicates whether there is a significant change of architecture for different experimental groups. This interaction was significant for Participants 2, 3 and 5, therefore indicating a possible change of architecture, due to the configural changes in the face.

From Figure 30 it can be observed that, across all participants, there are fewer errors in the old-face condition than for both the old-configural and new-configural faces. The main variable of interest for this study was RT, and the participants had a goal to reduce their error levels. The error rate was not analyzed separately, given that for some cases it was very small with no variance. We performed a repeated measures analysis by averaging the group error proportions across all participants ($F(2,10)= 7.87$, $p<0.009$; $M_{old\text{-}configuration}=.029$, S.E.Mean=.004, $M_{new\text{-}configuration}=.047$, S.E.Mean=.008, $M_{old\text{-}faces}=.012$, S.E.Mean =.002). So, the error level was, on average, lower than 5%, and consistent with the mean RT analysis.

Overall we conclude that there is a significant trend for all participants to process configurally altered faces slower than the old faces, for both face manipulations (the old-configural and old-standard faces). Both mean RT and mean error level replicates the findings from previous studies that revealed similar effects of configuration change (Tanaka & Sengco, 1997).

Comparison of the processing characteristic differences between old faces and new faces based on removing the face context (featural change) (Featural x Standard design)

The same GLM analyses were run on these trials. The only difference was in the data source. The group 1 observations (old configural faces) are collected from the part experiment where old whole faces are combined with the feature-only faces (see Figure 26), the group 2 data stems from the same subexperiment, but now consists of feature-only faces, and the group 3 data stems from the standard comparison experiments, with old whole faces, same as in the previous analyses. The goal of this experiment was to determine the effect of removing the face background, (which was not important for the correct decision), on mental organization during face processing. Although the face background was not necessary to make a correct decision, if face encoding is based on gestalt/holistic processes, we would expect that by removing it, we would have a larger effect of the configural/holistic strategies rather than the analytic.

The results of GLM analyses, comparable to the previous section, are presented in Table 4, right. Also, mean RT for the different experimental groups are presented in Figure 26 (old-featural and new-featural groups).

From the results presented in Table 4 (right side of the table) we can see that all the main effects of Lips and Eyes were significant at the $p<0.01$ level with power=1. Also, the main effect of experimental group was significant for all participants, which indicates that there is a significant mean RT difference between experimental groups. In Figure 4 we present mean RTs for the different experimental group: All participants exhibited a similar pattern of mean RTs: $RT_{old\text{-}faces} \approx RT_{old\text{-}feautres} < RT_{new\text{-}configuration,}$ when old faces are the whole faces from the standard condition, old-featural faces are the old faces but presented together with featural faces, and new-configuration are the new featural-faces.

Also, the MIC tests revealed by the two-way interaction Lips x Eyes were significant for all participants with very high power=1. Overall, MIC values were positive for all participants, revealing overadditivity. However, in further analysis, the MIC values will be broken down for different experimental conditions, and for each participant (see Figures 31, 32 and 33).

The three way interaction Eyes x Lips x Exp group indicates whether there is possibly a significant change of architecture for different experimental groups. This interaction was significant for most participants, with Participants 3 and 4 reaching marginal significance and with the exception of Participant 6.

From Figure 30 it could be observed that, across all participants, there is a clear trend of higher error reduction in both the old-faces and old-featural conditions than for the new-featural faces. We performed a repeated measures analysis by averaging the group error proportions across all participants ($F(1.013,5.063)= 6.049$, $p<0.056$), which was of marginal significance ($M_{old\text{-}featural}=0.014$, S.E.Mean=0.0017, $M_{new\text{-}featureal}=0.070$, S.E.Mean=0.023, $M_{old\text{-}faces}=0.012$, S.E.Mean=0.002). Note that we used the Greenhouse-Geisser test because the sphericity assumption was violated. Thus, the error level was on average, lower than 7%.

Overall we conclude that there is a significant trend for all participants to process featurally based faces both slower and with more errors than the old faces.

Mean and Survivor Interaction Contrast Functions

Both the mean interaction contrast (MIC) and survivor interaction contrast test results are presented in Figures 31, 32 and 33, for all experimental groups of interest. The results are only for the gang members, because the design only allowed their data to be tested for architecture. We present data from each subexperiment.

We focus on the SIC functions presented in Figures 31, 32 and 33. Overall, the data are dominated by positive unimodal SIC functions for different subexperiments and participants. The positive SIC function with one peak is a strong indicator of parallel self-terminating (minimum time) processing architecture. Also, this finding is supported by the appropriate ordering of survivor functions (first column in Figures 31, 32 and 33) combined with the overadditive MIC results. With the exception of several cases within the feature test, primarily the ordering of survivor functions is persevered.

In the standard test (Figure 31) we can see that all participants exhibited clear, positive SIC functions, while their MIC values revealed overadditivity. Further, Participants 2 and 6 exhibited small negative deviations that may be indicators of a coactive processing structure. We conclude that presentation of old faces in the OR task utilized parallel processing architecture on two processing features, and participants could terminate processing as soon as one feature was recognized.

The configural-test group revealed similar results at the SIC and MIC levels, with appropriate ordering of survivor functions (Figure 32). We conclude that when faces are configurally altered, that is when their background is removed and replaced with a new one, the processing architecture of the face features remained parallel self-terminating. Note that although Participant 4 exhibited a small positive MIC value, only one face dimension (the eye separation) was significant. However, Participant 4 (Table 3) also exhibited the fastest mean RT in the feature-test condition in contrast to the configural-test and standard-test conditions, for the gang-member faces. We suggest that this participant qualitatively changed processing strategy which was based on a more analytic than holistic approach.

The feature-test SIC analysis revealed a different picture: several Participants (1, 2 and 5) did not exhibit regular MIC and SIC functions (Figure 33). In fact it is evident that they showed significance of the eye separation only, while no effect was delivered from the lip position. The other three participants showed regular parallel self-terminating SIC functions, which was also supported at the MIC level. However, these findings are not surprising. Without the face-background face features loose their support system and their spatial tags. While the eye separation can work as an independent feature because it is possible to learn the spatial relation between them, the lips are more detrimentally affected because they lost their immediate points of reference and can now only be related to the eyes.

We averaged the MIC scores over all participants and ran a repeated measures GLM analysis (Table 5). The main effect of the difference between averaged MICs was not significant, but note that the number of participants is rather small. Interestingly, the standard test exhibited a lower mean MIC score than the configural test, while the featural test revealed the smallest MIC value. We suggest that a large part of the magnitude of MIC value is due to the presence of whole face structure regardless of whether it is old (standard) or new (configural).

Table 5: Mean MIC values and their standard deviations, averaged over all participants, for three subexperiments, for gang-members faces.

	Mean MIC (ms)	Std. Deviation
Standard-test	82.4341	37.61489
Configural-test	128.8395	94.14451
Feature-test	39.4559	44.60154

It is interesting to note that ordering at the mean RT level for almost all participants was $RT_{standard\text{-}test} < RT_{feature\text{-}test} < RT_{configural\text{-}test}$, for the gang-member faces. It could be expected then that both the SIC and MIC would show some monotonic transition between conditions, such as diminishing or increasing magnitudes. But when compared between groups, the SIC functions did not show a transition. The second very interesting aspect of the feature test is that, on average, attenuation of one face feature (lips) produced faster processing when compared to the introduction of a new face background (the configural test). So we can reconstruct the following scenario: removing the face background attenuated one face dimension and produced slower processing on average than with the standard old faces. Now, putting on a new face background slowed processing even more, probably because the new face background provided invalid reference cues for detection of spatial relations between features which could be realized through dependency between face units and, at the same time, re-engaged processing of the previously attenuated lips, producing overall significant effect on them.

If the last is true then the capacity test should reveal the smallest CCF for the featural-test condition when compared with both the standard-test and configural-test conditions.

<u>The capacity test</u>

We calculated several different capacity functions for each participant. In Figures 34 and 35 we present the calculated capacity functions, along with bootstrapped 90% percentile confidence intervals, for both the gang-member faces (pooled together into one condition) and the hero face. So for each participant we calculated 8 capacity functions, 4 for the gang members and 4 for the hero face, in the blocked conditions.

We expect that the holistic advantage of processing old whole faces will be evident as super capacity processing, i.e. the capacity coefficient function confidence interval bounds will swing above the $C(t)=1$ value. Subsequently, we expect that the CCF confidence interval for the new configural-faces test will be below $C(t)=1$ value. We assume that new configuration will impose lot of processing demands because the cognitive system has to process unknown global face configuration that has not been observed before, and probably slower processing can be engaged (serial). The weaker assumption of a holistic/configural effect does not necessarily predict super capacity, but a significant magnitude difference between CCFs for old faces and new faces.

Also given that we found that processing of old faces is affected by mixing them with new-configural faces, we expect that in that the CCF for the standard test old-faces to be bigger then the CCF for old faces configural test.

As far as the feature-test condition is concerned, we expect that the feature faces (from the featural-test condition) will exhibit CCFs that are below $C(t)\leq 1$, thereby revealing unlimited or limited capacity. In a weaker form, this assumption predicts that it is not necessary for the value to be lower than 1, because we can expect some configural properties to emerge even from the feature only presentation, given that some configuration can be inferred from the presentation of the eyes and lips only. So in this weaker form, it is possible that the CCF for feature-only faces exhibits even moderate upper violation of $C(t)=1$ value, toward super capacity.

By comparing the CCFs obtained for different conditions, we expect that any holistic/configural processing will be evident in strong violations of $C(t)=1$ value for the old faces. We expect limited capacity for both new faces and for feature faces.

The weaker assumption does not assume violation of some value of C(t), but that there should be an ordering effect between old whole faces and new faces, such that old-face CCF is significantly bigger than the new-face and feature-face CCFs.

We also predict that the hero face should exhibit a similar pattern of results to the gang-faces, for different conditions.

Now we will introduce two methods for calculating CCFs. Note again that the goal is to calculate a measure of gestalt/holistic properties of whole face encoding. This is reflected in the amount of work done on whole faces, compared to the cognitive effort needless to process part-based faces. So, this function is a relative measure, and it could be affected by the learning of the part-based information, if any occurs. It is debatable whether a feature or part-based information could be learned as a gestalt or could posses a good form. With respect to the amount of learning of the part-based information, we can derive two CCFs. In the **first method**, the part-based trials of the old faces used for calculation, stem from the actual experimental sessions. Implicitly, we assume that there is no additional learning conducted on part-based information. In the **second method**, the part-based trials are from the very first learning session, from the initial phase of the study. Note that here we are dealing only with the learning aspect of the part-based information from the old faces, not the new faces.

Method 1

The CCFs were calculated by taking the ratio of integrated hazard functions, which is a measure of the amount of work done on old whole faces, and the amount of work done on the part-based faces. We used the trials only from the standard test for the numerator (the old whole-faces) trials, and so are taken from the first part of experiment where whole faces were not combined with part faces. On the other hand trials for the denominator of the CCF, or the part-based faces, stem from the capacity test, which is taken from the mixed condition.

Method 2

In the second method, the CCFs are calculated by changing the source for the denominator origin. In fact, the trials for the denominator are taken from the very first learning sessions. The rationale for this was that if we are to compare the effects of encoding of whole old faces to the whole new faces (configurally changed), then we should use the denominators that reflect the same amount of learning of part-based faces as the whole face learning portion (numerator). For the old faces, these are very first trials of the learning sessions, while for the new configural faces, the part-based faces were introduced for the first time in the test phase. So in order to get a better description and eliminate different levels of learning of part-based old and new faces, we have to use trials of part based faces that equate the degree of learning.

The capacity test results

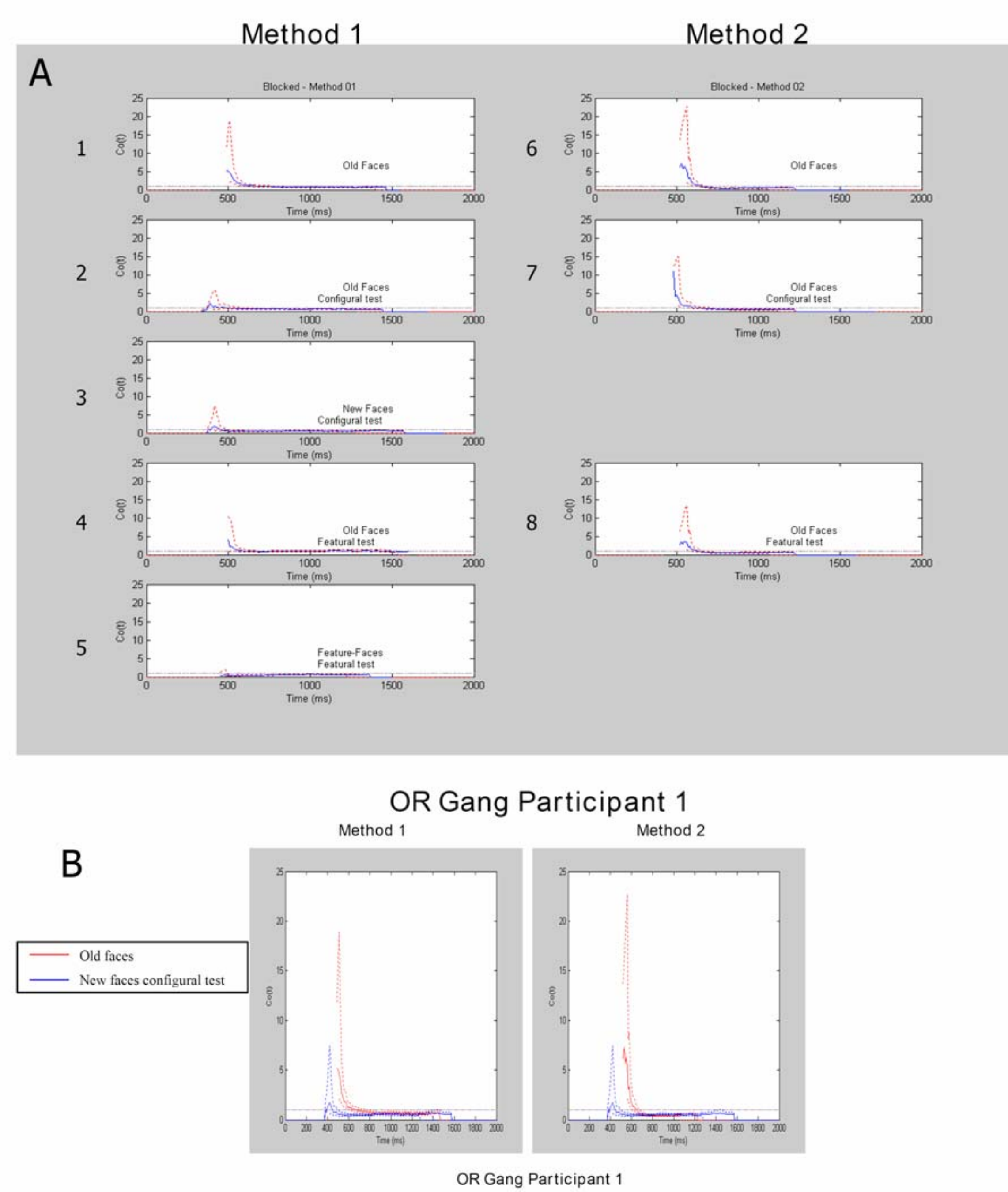
We present the calculated CCFs for different experimental conditions for the gang-members, in Figure 34, for each participant. Both old and new whole face-conditions were taken from the blocked sessions, where they were not mixed with part-based information. The first function (top of Figure 34 A) represents the CCF for the old faces from the standard-test condition, where only old faces were presented alone. The next function for the old-faces configural test also shows the CCF for the old faces, but taken from the experimental condition when they were presented together with the new configural faces. The new faces configural-test represents the CCF for new faces from the same test session. The old faces featural-test represents the CCF when old faces are presented together with feature-only faces in the session (see design figure 26, the 2nd part, feature-face test), and the feature faces from the featural-test function correspond to the CCF for feature faces only from that same condition. The right column of Figure 34 B shows the same CCFs calculated by the second method.

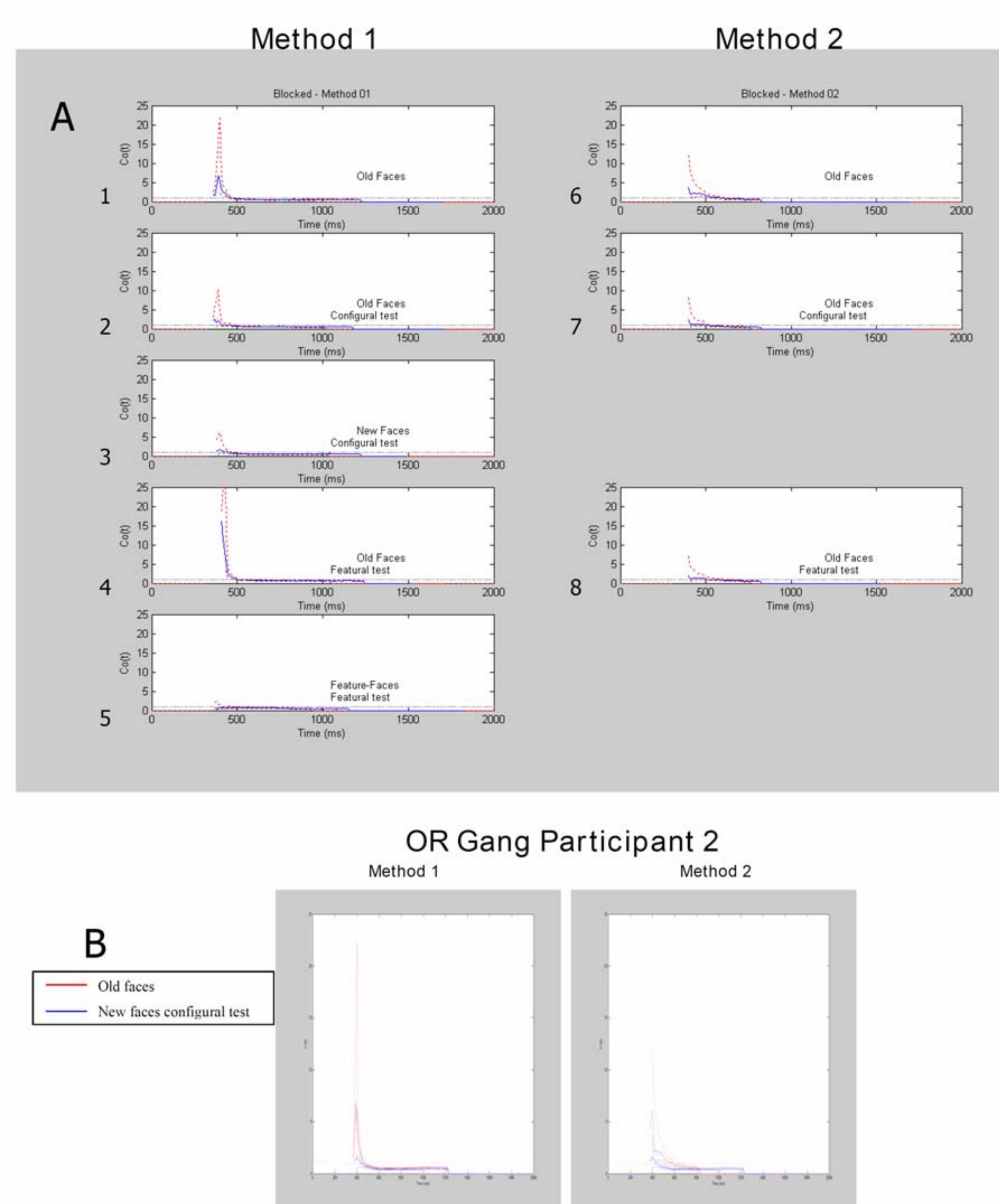
Overall, we observe that for the most participants, the old-faces CCF exhibited the highest value and dominates the other conditions. Also, we can observe that the weaker assumption of configural effects is satisfied for almost all cases; that is, $CCF_{old\ faces} > CCF_{new\ faces}$. Second, for almost all participants, the CCFs for the new configural faces include C(t)=1 in the confidence intervals and therefore suggesting the absence of super capacity. These values of CCFs do not suggest that the configurally altered faces exhibited limited capacity processing. When feature faces were, processed CCFs are below C(t)=1, for most of the time and are generally of lower magnitude than for the new faces. This implies limited capacity of the features only.

We will now perform separate analyses on the gang-members and the hero face.

Figure 34: (A) calculated CCFs for different experimental conditions for gang-members, for each participant, presented on a separate page. Both old and new whole face-conditions were taken from the blocked sessions, where they were not mixed with part-based information. (1) represents the CCF for the old faces from the standard-test condition, where only old faces were presented alone. (2) CCF for the old faces configural test, but taken from the experimental condition when they were presented together with the new configural faces. (3) new faces configural-test represents the CCF for new faces from the same test session as the second CCF. (4) The old faces featural-test represents the CCF when old faces are presented together with feature-only faces in the session. (5) CCF for the feature faces from the featural-test function feature faces only, from that same session as the fourth CCF. The (6), (7) and (8) CCFs are calculated by the second method.

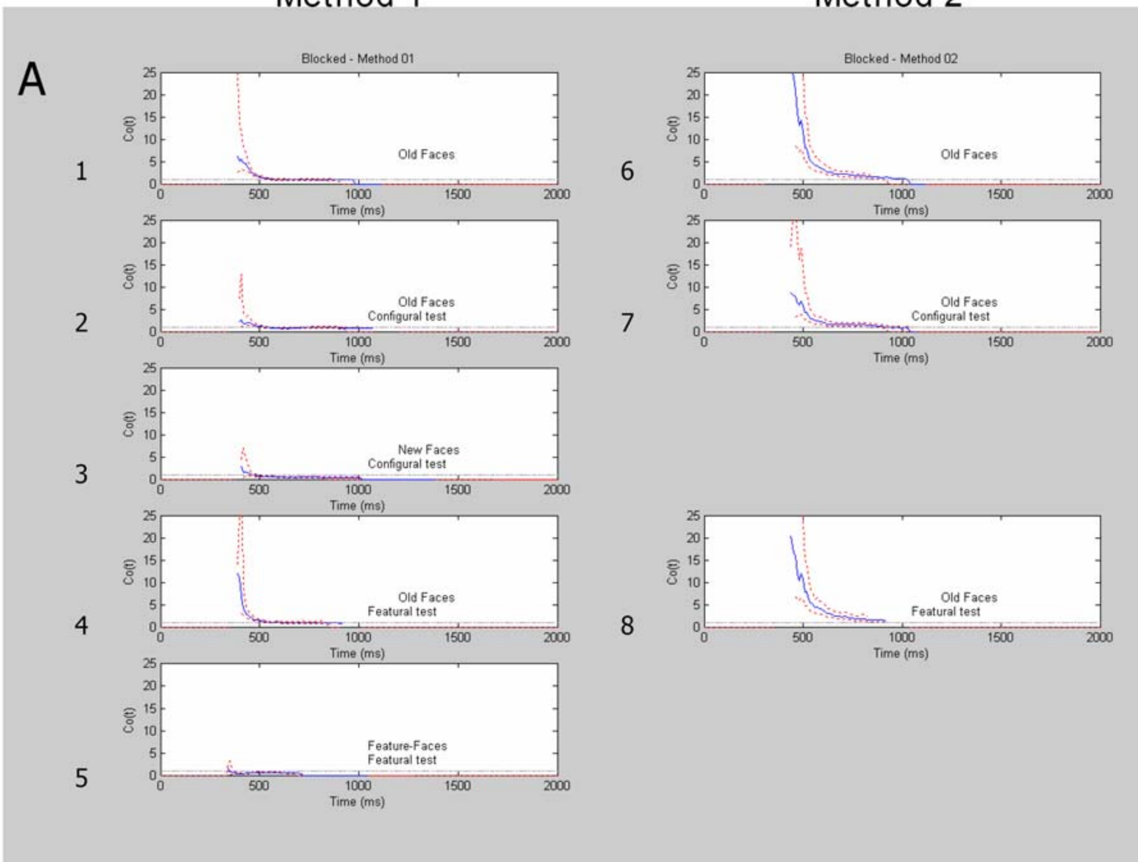
(B) The configural test based on a comparison of bootstrapped confidence intervals, between old faces CCFs (calculated by two methods) and new faces CCFs. Figure on the left is a comparison between (1) and (3) CCFs; and figure on the right is a comparison between (3) and (6) CCFs.


Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Old Faces
Old Faces
Configural test
New Faces
Configural test
Old Faces
Featural test
Feature-Faces
Featural test
Time (ms)
Co(t)
OR Gang Participant 1
B
Method 1
Method 2
Old faces
New faces configural test
OR Gang Participant 1



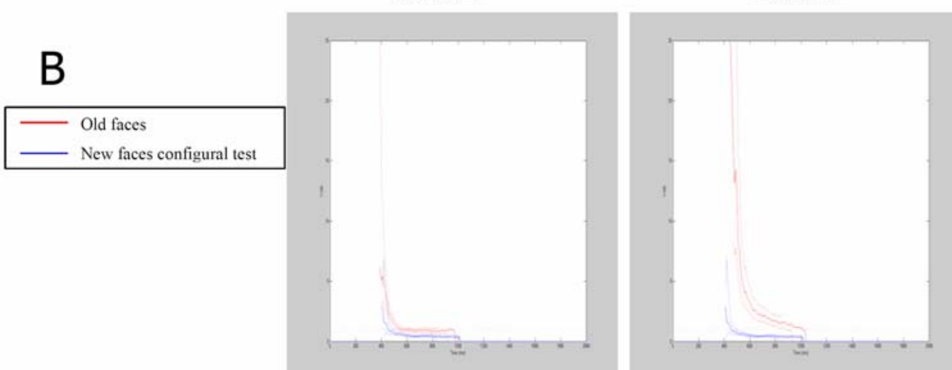
OR Gang Participant 2

Method 1 Method 2

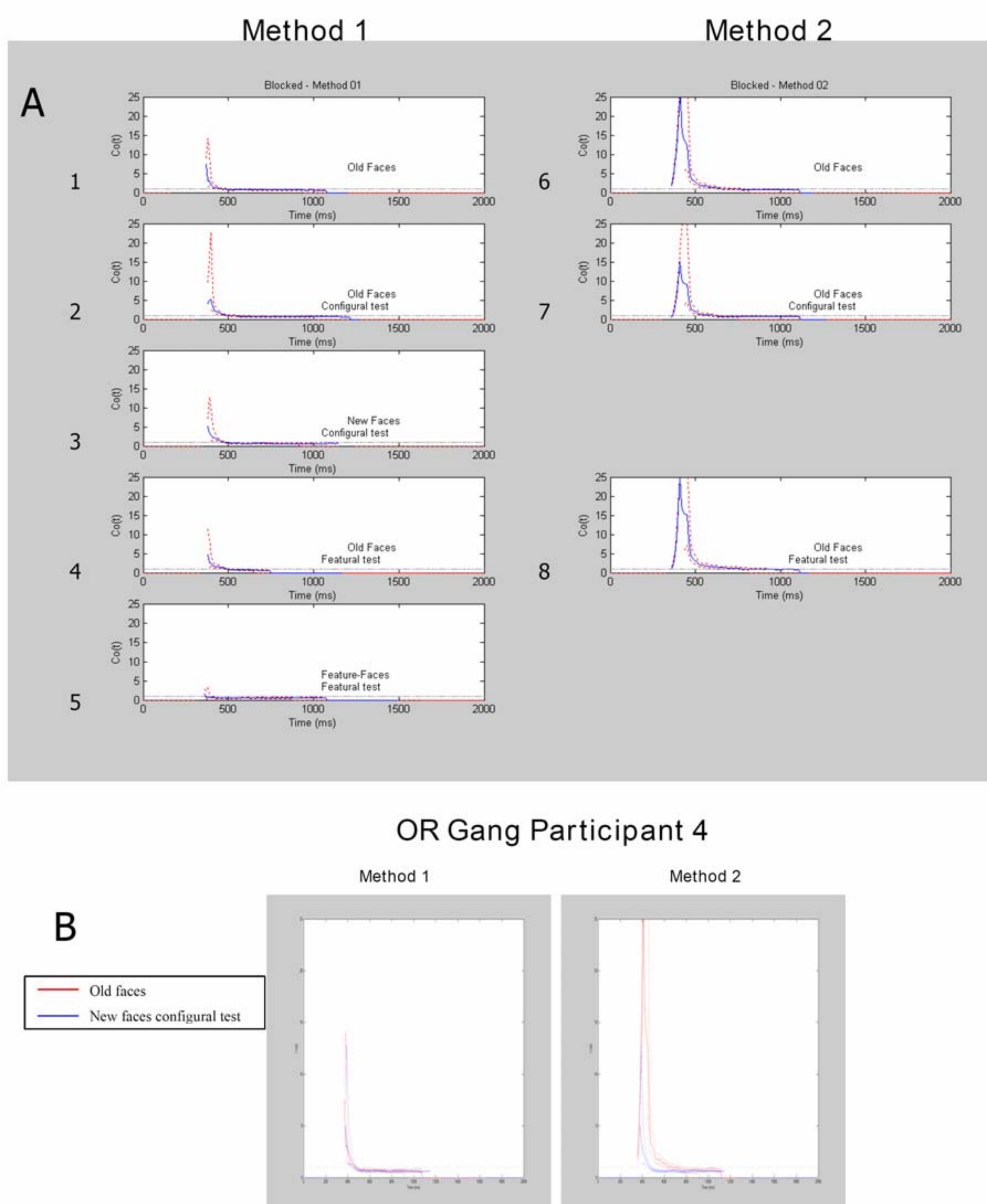


OR Gang Participant 3

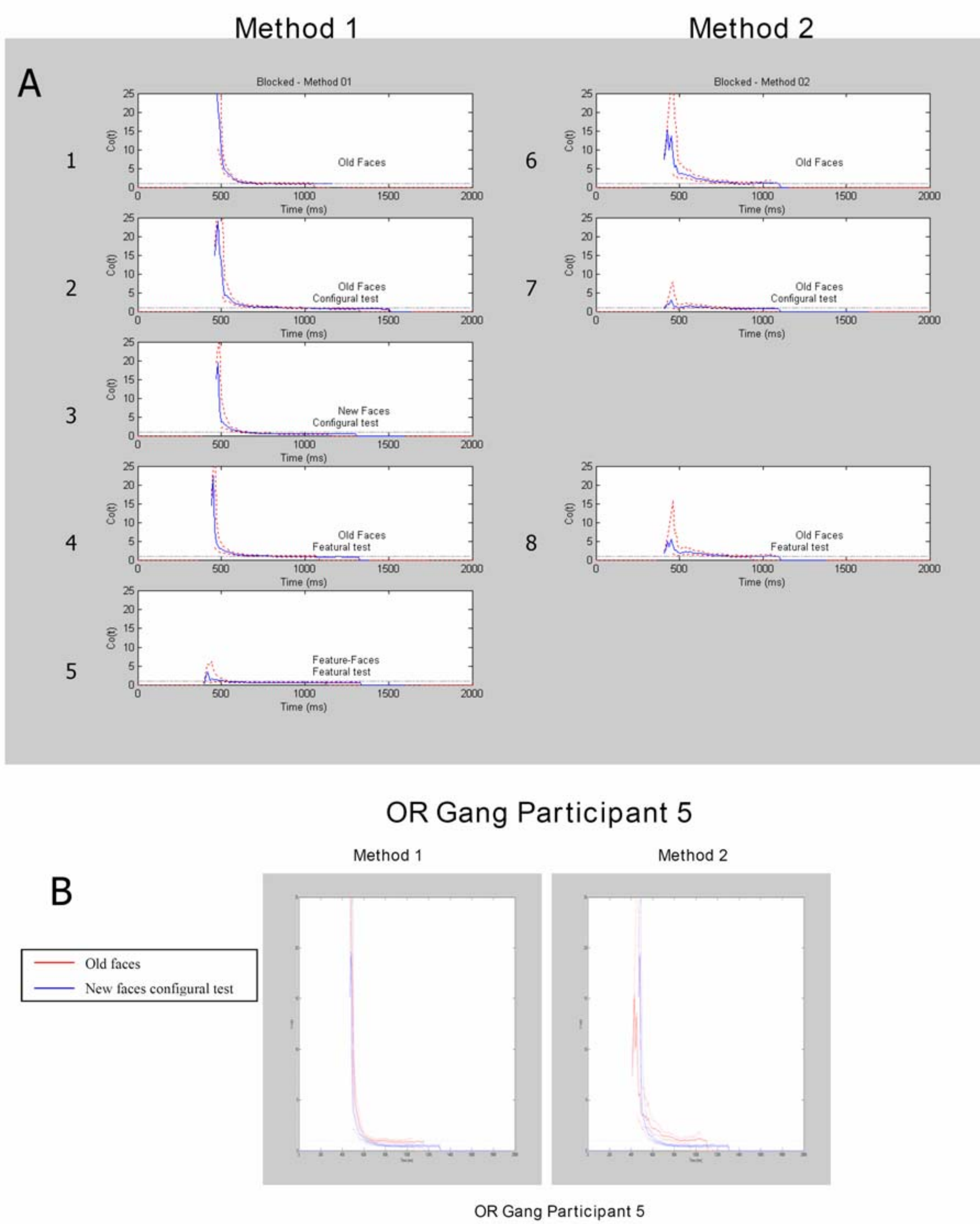
Method 1 Method 2



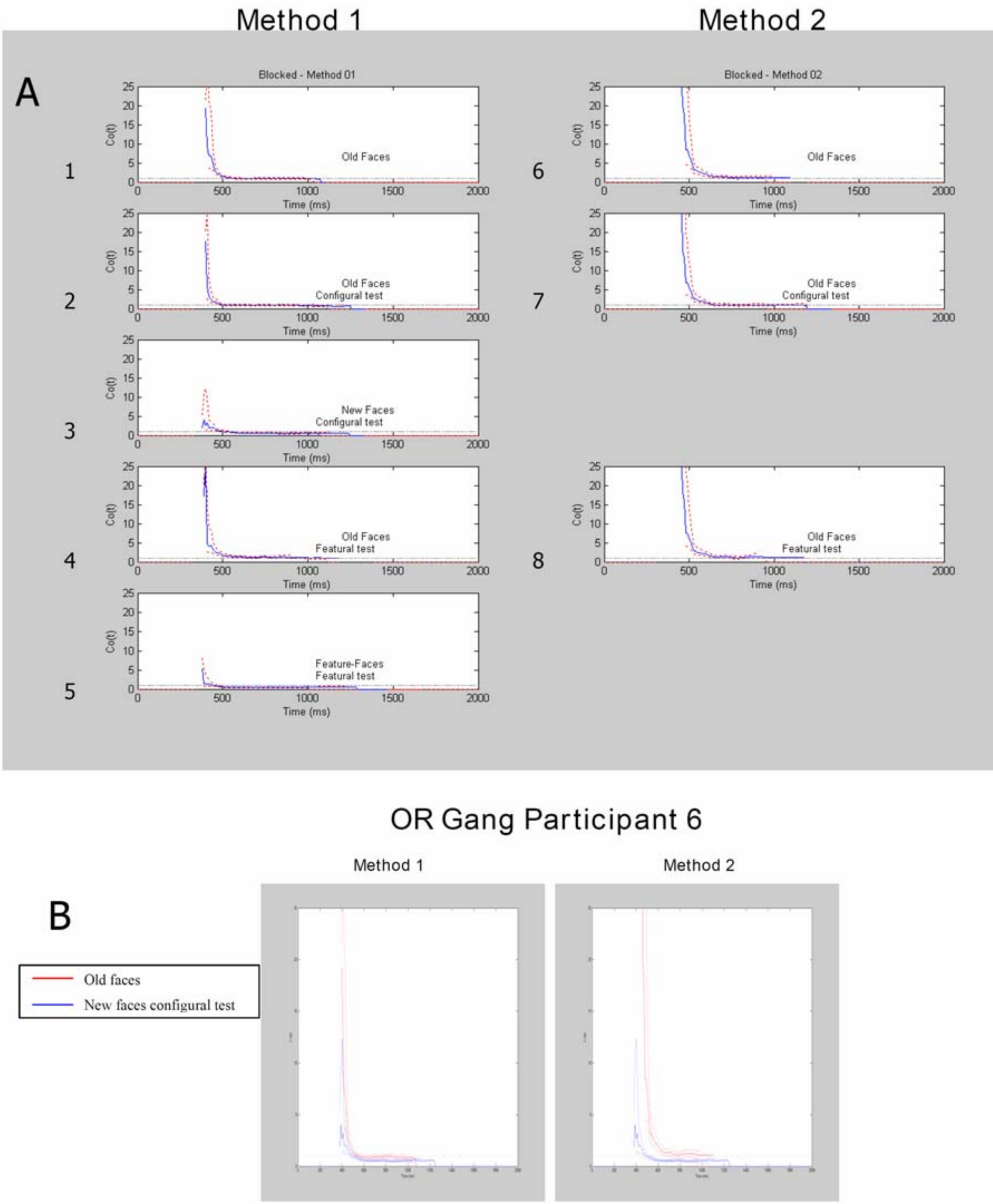
OR Gang Participant 3



OR Gang Participant 4



OR Gang Participant 5



OR Gang Participant 6

The configural effect and capacity functions for the gang faces

Figure 34 consists of multiple graphs, for a single participant. Overall, we can observe that all old faces from the standard test (first plot) calculated by both methods exhibited super capacity for all participants. New faces exhibited unlimited capacity for Participants 1, 2 and 3, while Participants 4, 5 and 6 exhibited super capacity. The feature faces for all participants consistently exhibited unlimited to very limited capacity. Not all old faces in both the configural-test and featural-test were super capacity, which is consistent with our previous finding that, on average, processing is slower when old-faces are combined with altered faces in the same experiment.

On another figure we report two graphs of the most interest: the comparison between old faces (calculated by two methods) and new faces. This test is based on a comparison of the magnitude of CCFs and their corresponding confidence intervals. In order for the two CCFs to exhibit a significant difference, both functions must be separated from each other that all confidence intervals do not overlap, at least at some point in time. We also expect that the difference if it exists will follow a monotonic relationship and that no change of sign of that difference will be observed at any time. That is, we expect there to be no reversal of the effects of the expected difference.

When compared for the configural effect, Figure 34 B, all participants exhibited a significant difference between the old face and new face CCFs for both calculation methods. We found that in all cases the old face CCF was statistically larger in magnitude than the new face CCF. The exceptions to this were Participants 2 and 4, but only for the method 1. We conclude that configurally altering the old face into the new face resulted in a reduction of CCF magnitude, which could correspond to a reduction in the dependency between face features. Half of the participants exhibited super capacity when processing the new faces, so the introduction of a novel facial surround was not very detrimental to processing, as we expected.

On the other hand, removing the face background (the featural-test) never produced super capacity, but resulted in a combination of both unlimited and limited capacity, with predominantly limited capacity processing.

The configural effect and capacity functions for the hero face

The following results are very similar to the processing of the gang-member faces.

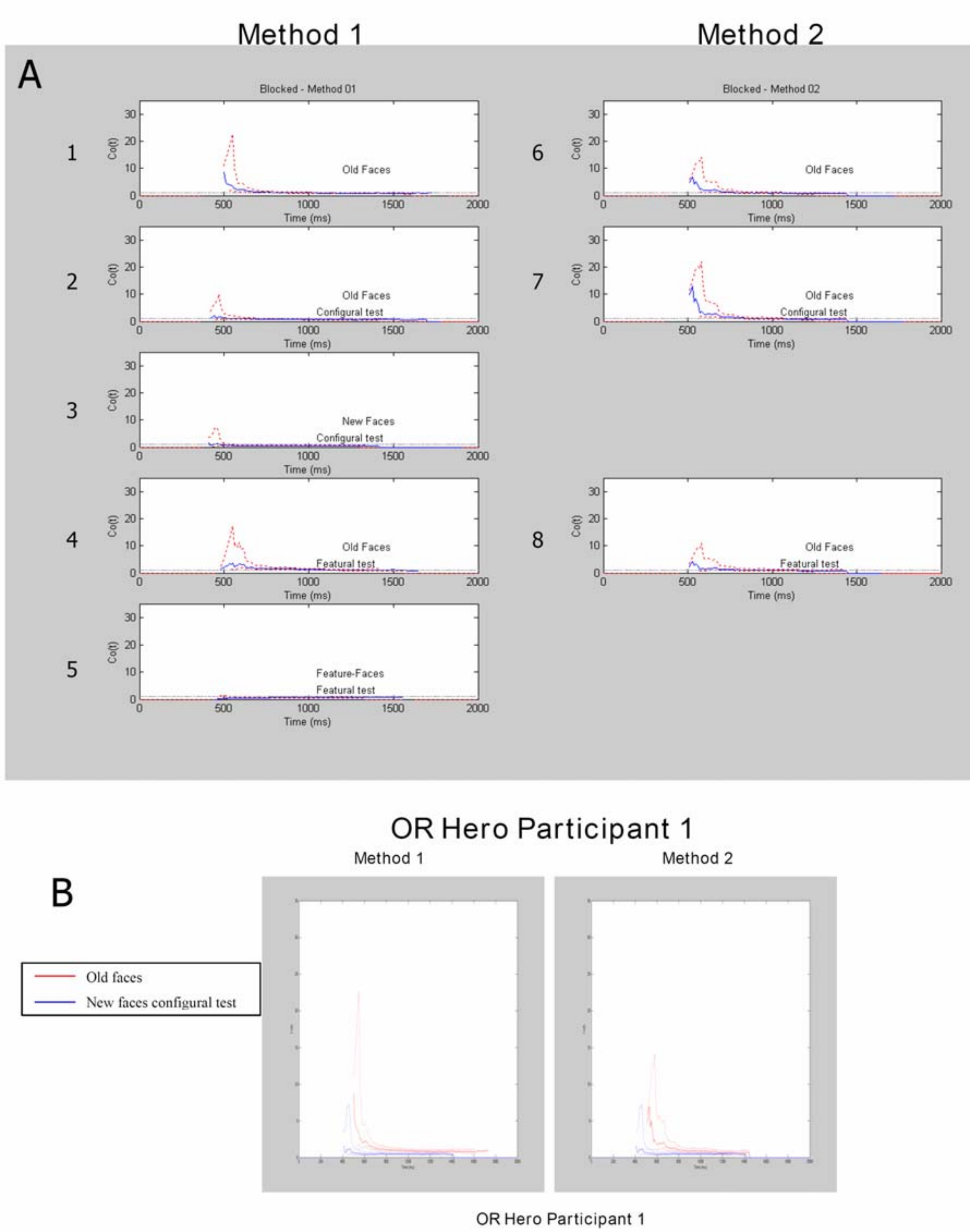
Figure 35 consists of multiple graphs as described above, for a single participant. Overall, we can observe that all old faces from the standard test (first plot) calculated by both methods exhibited super capacity for all participants.

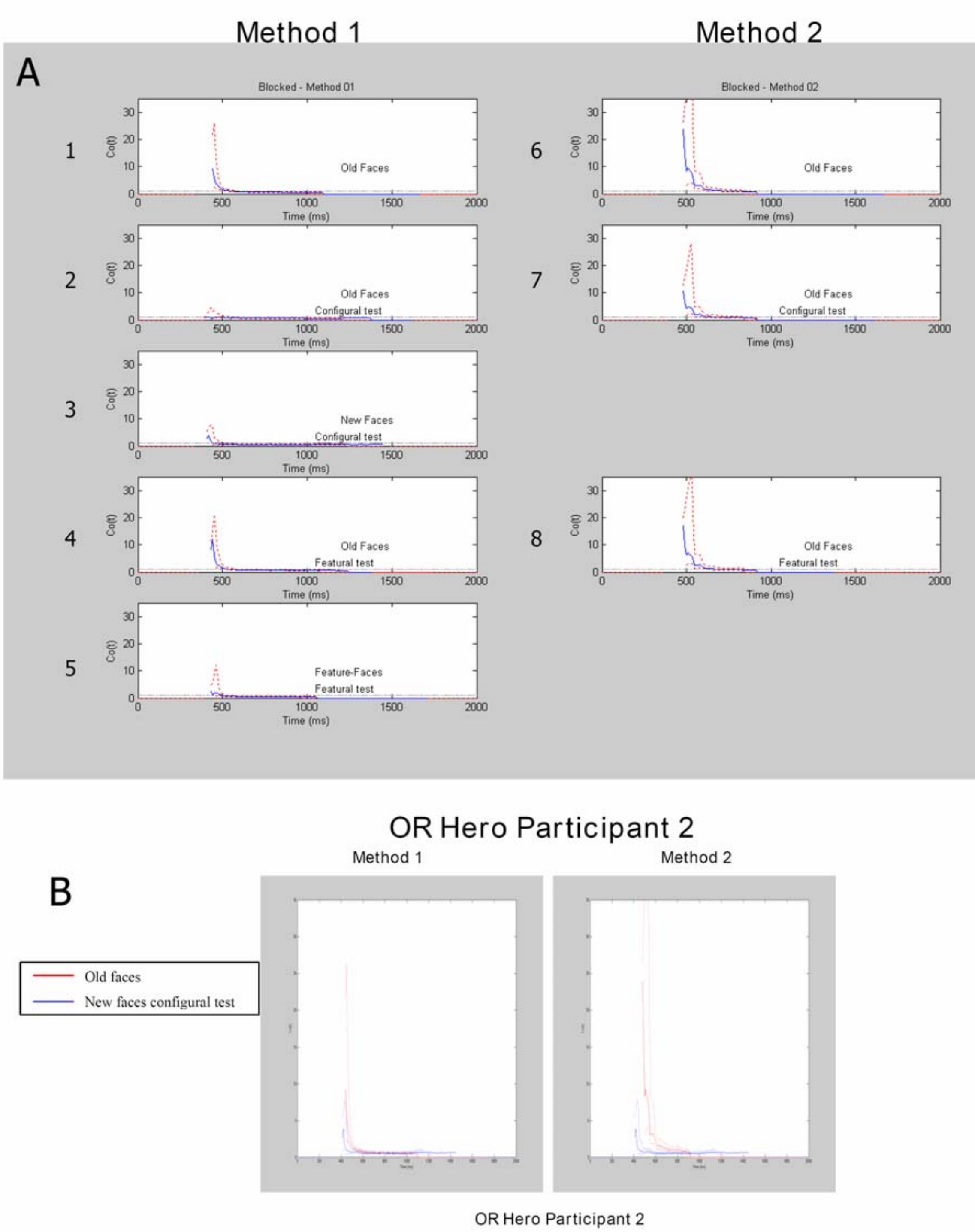
New faces exhibited unlimited capacity for Participants 1, 2, 3 and 6, while Participants 4 and 5 exhibited super capacity. The feature faces for all participants consistently exhibited unlimited to very limited capacity. Not all old faces in the configural-test were super capacity (Participants 1, 2 and 5). This is consistent with our previous finding that, on average, processing is slower when old faces are combined with altered faces in the same experiment. All old faces in the featural-test were super capacity.

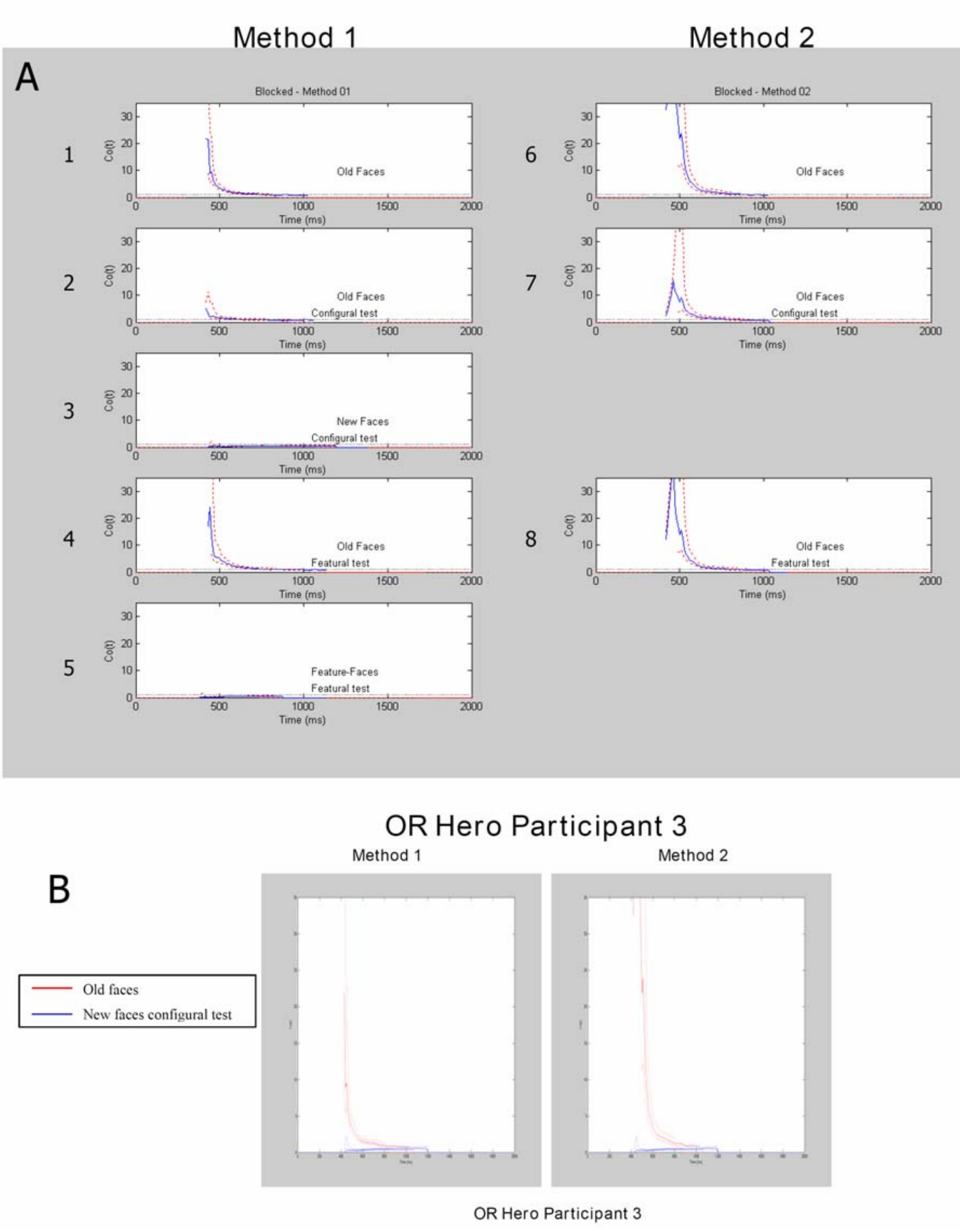
When compared for the configural effect, Figure 35 B all participants exhibited a significant difference between the old-face and new face CCFs for both methods. The old faces exhibited a CCF with statistically larger magnitude than new face CCF. Two participants (4 and 5) exhibited super capacity when processing the new faces, so the introduction a novel background was not very detrimental to processing in those cases. Removal of the face background (the featural-test) never produced super capacity but resulted in a combination of unlimited and limited capacity, with more prominent limited capacity.

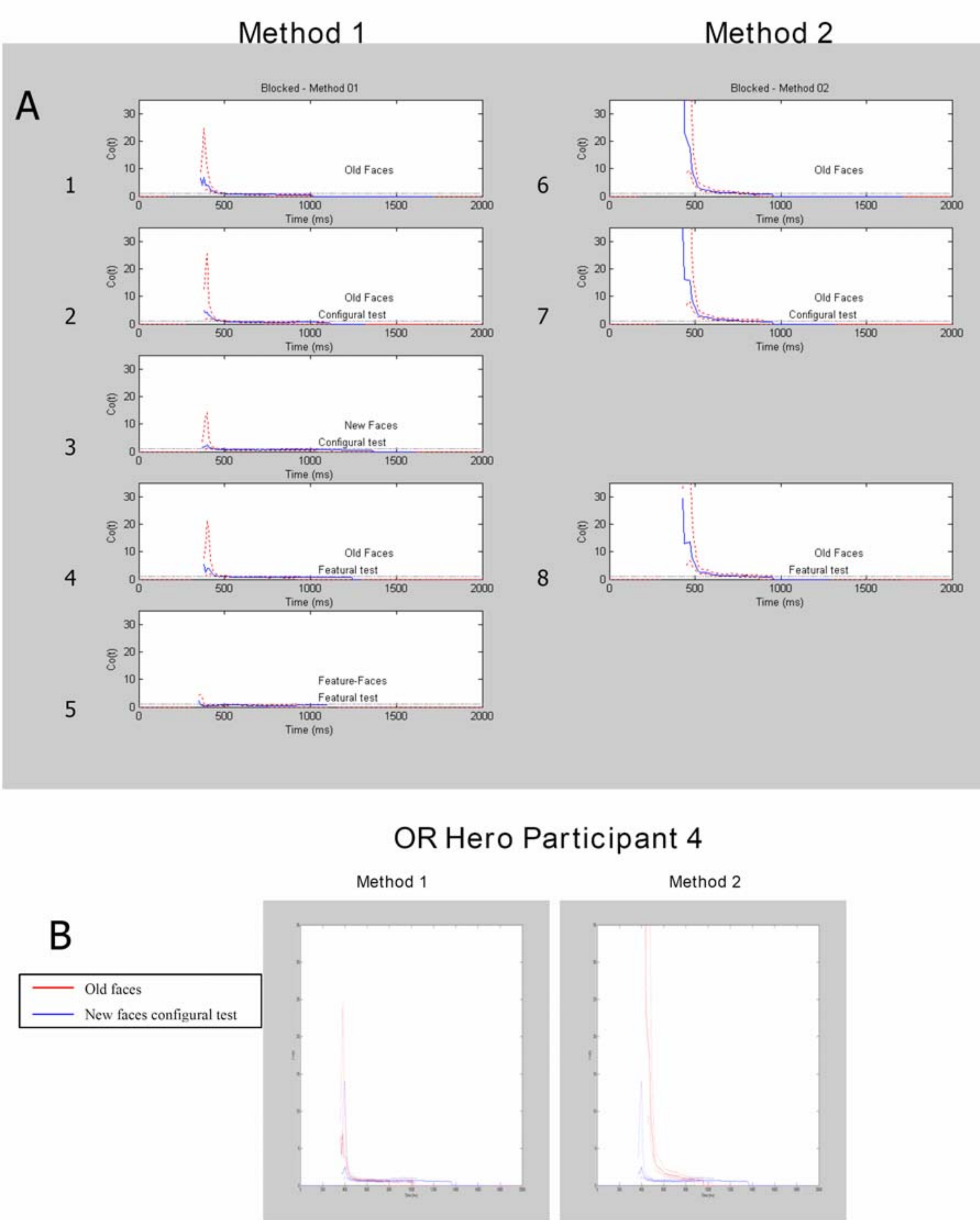
Figure 35: (A) calculated CCFs for different experimental conditions for hero face, for each participant, presented on a separate page. Both old and new whole face-conditions were taken from the blocked sessions, where they were not mixed with part-based information. (1) Represents the CCF for the old faces from the standard-test condition, where only old faces were presented alone. (2) CCF for the old faces configural test, but taken from the experimental condition when they were presented together with the new configural faces. (3) New faces configural-test represents the CCF for new faces from the same test session as the second CCF. (4) The old faces featural-test represents the CCF when old faces are presented together with feature-only faces in the session. (5) CCF for the feature faces from the featural-test function feature faces only, from that same session as the fourth CCF. The (6), (7) and (8) CCFs are calculated by the second method.

(B) The configural test based on a comparison of bootstrapped confidence intervals, between old faces CCFs (calculated by two methods) and new faces CCFs. Figure on the left is a comparison between (1) and (3) CCFs; and figure on the right is a comparison between (3) and (6) CCFs.

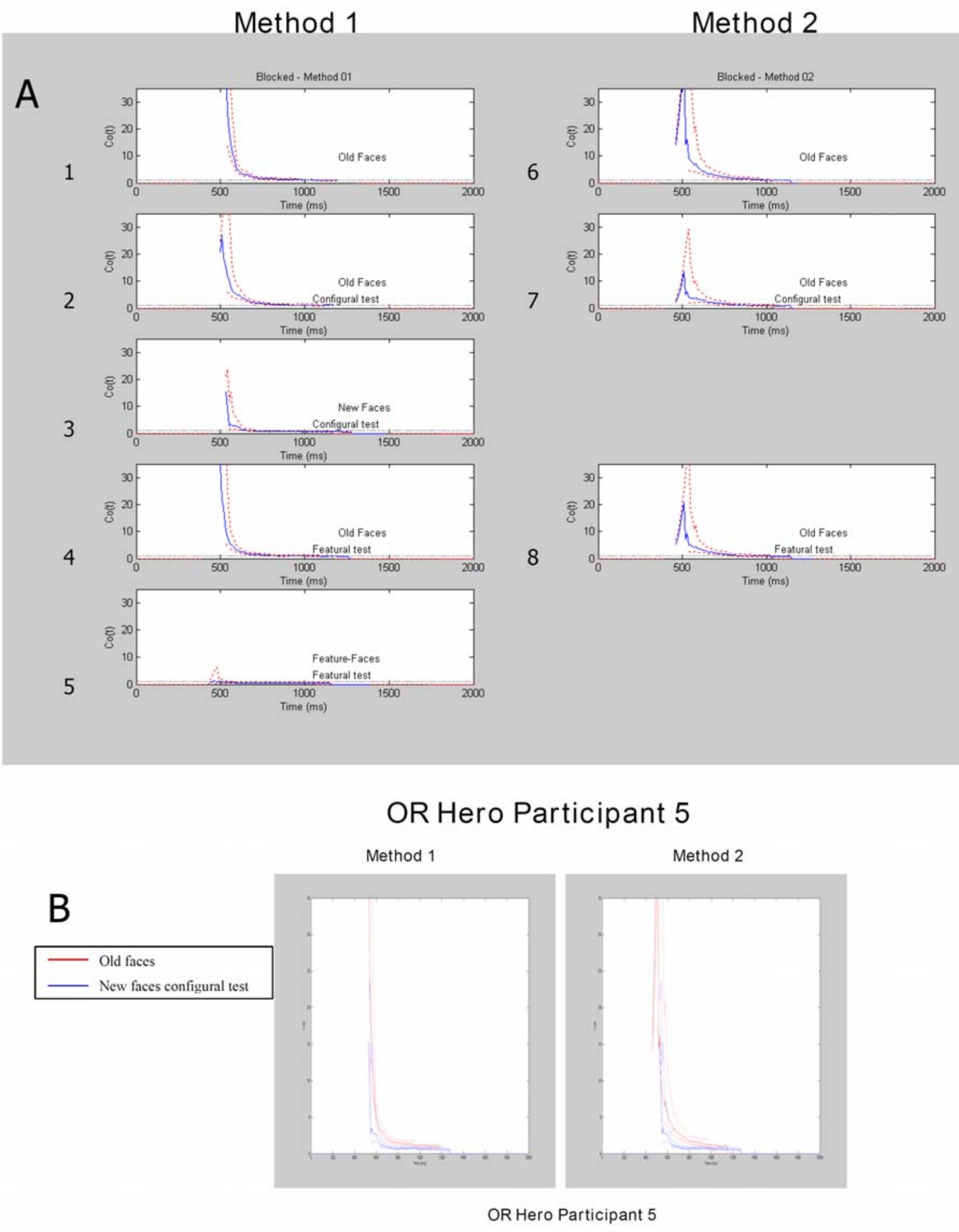

Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Old Faces
Configural test
New Faces
Featural test
Feature-Faces
Co(t)
Time (ms)
B
OR Hero Participant 1
Old faces
New faces configural test
OR Hero Participant 1


Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Co(t)
Time (ms)
Old Faces
Old Faces
Configural test
New Faces
Configural test
Old Faces
Featural test
Feature-Faces
Featural test
OR Hero Participant 2
B
Old faces
New faces configural test
OR Hero Participant 2

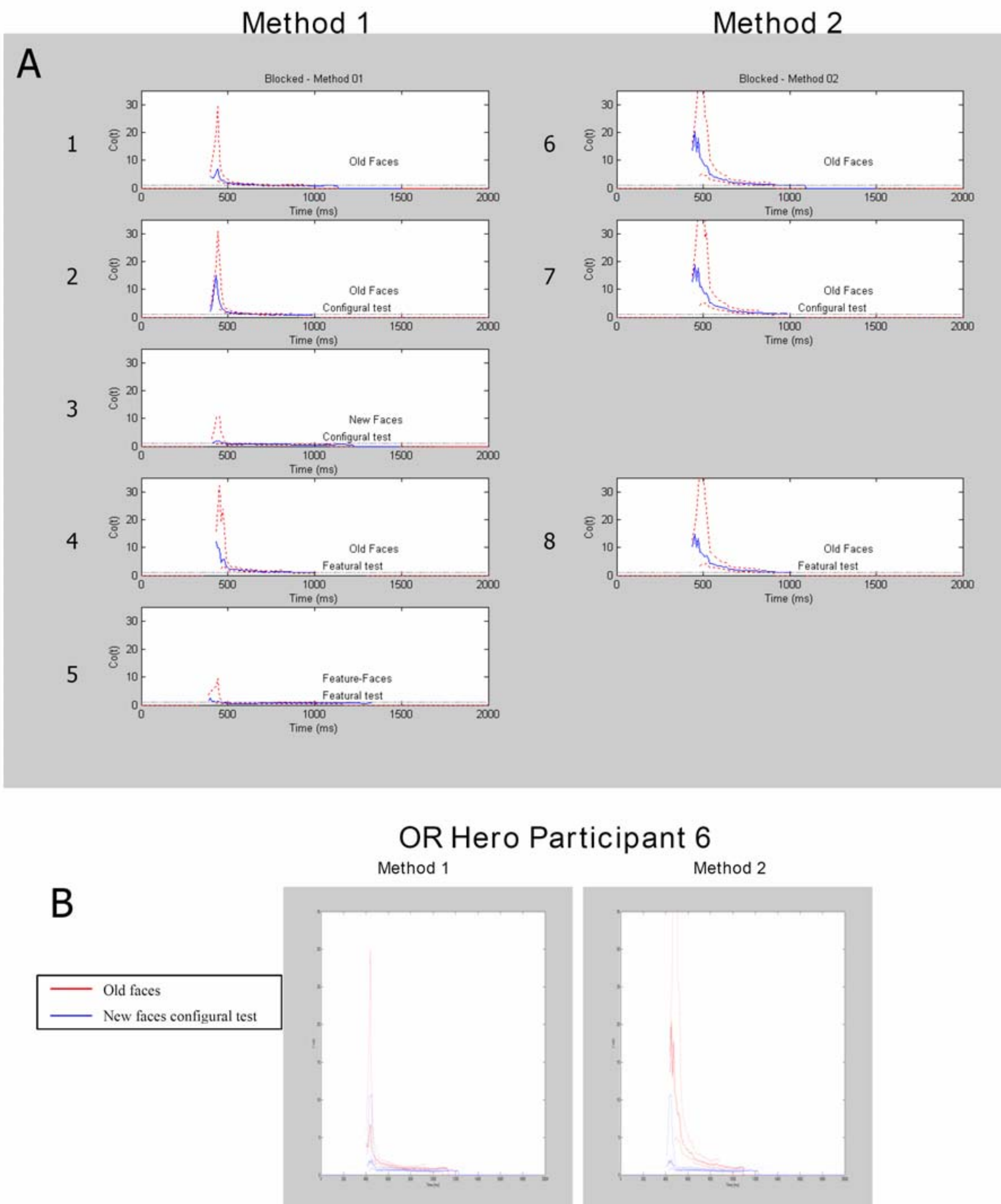

Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Co(t)
Time (ms)
Old Faces
Old Faces
Configural test
New Faces
Configural test
Old Faces
Featural test
Feature-Faces
Featural test
OR Hero Participant 3
Method 1
Method 2
B
Old faces
New faces configural test
OR Hero Participant 3



OR Hero Participant 4



Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
Co(t)
Time (ms)
1
2
3
4
5
6
7
8
Old Faces
Old Faces Configural test
New Faces Configural test
Old Faces Featural test
Feature-Faces Featural test
OR Hero Participant 5
B
Old faces
New faces configural test
OR Hero Participant 5



OR Hero Participant 6

We conclude that in the OR case, face background plays an important role in face recognition/categorization and even if it was not analogous to the background that had been previously learnt.

AND condition

Experiment learning phase

The same design and analyses have been applied in the AND condition

Method

Participants

Six participants, 4 females and 2 males were paid for their participation.

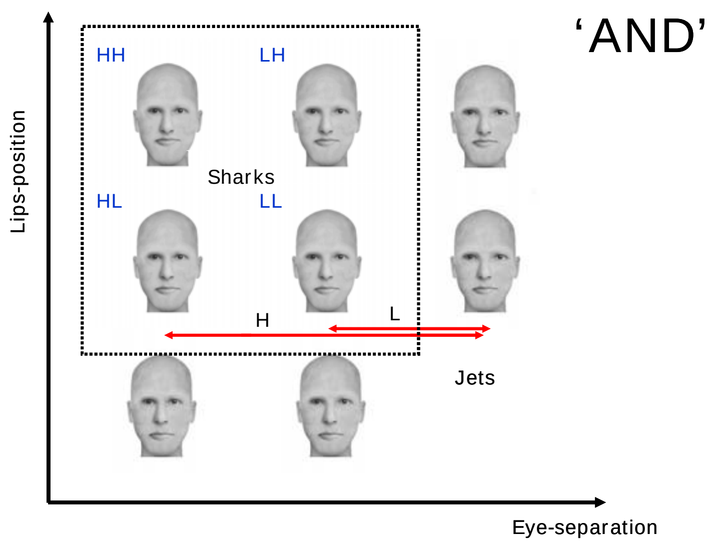
Materials

In the AND experiment, participants had to make a decision whether a displayed face belonged to one of two groups of gang-member faces: Jets or Sharks. In the first part of each session, participants observed only whole faces. The goal of this part was to investigate the organization of mental processing by application of the SFT tests (the MIC and SIC). In the second part, participants observed both whole faces and masked faces mixed together. The goal of the second part was to test the capacity of the system (the capacity test). In the third part, participants moved to a complete identification task, and they had to learn to identify 4 different faces, with associated names. The goal of this part was to implicitly let participants develop familiarity with some face properties that will be used in the test phase of experiment.

In the first part we manipulated the following factors: (a) face category, two categories were Jets or Sharks. Four member faces were designed for each gang; (b) feature configuration and (c) feature saliency.

Feature configuration involved changes to either the distance between the eyes or the height of the mouth relative to the nose. The saliency of features is defined by the marginal proximity of each face-feature projection onto the two-dimensional face space (see Figure 36) with respect to the other category group members. In this case the projection, or marginal value, of each face-feature for each gang-member is either close or distant with respect to the projection of same feature from the most similar member of the other gang. So (b) and (c) were factorially combined, in the each gang group. For the Sharks, factorial combination of the face features and the salience level produces four combinations: HH, HL, LH and LL. The first letter denotes the saliency level of the eye-separation, while the second letter denotes the saliency level of the lips position. So the condition HH defines one Shark gang member who is the most distant from the closest member of the Jet gang, in the designed face space. Here, the salience level H (high) results in faster recognition than the L condition. The saliency is connected to the similarity as well: low salience feature of one gang member face is more similar to the corresponding features from the other gang than high salience features. No control over the stopping rule could be imposed to the Jets members. In their case, termination of processing could occur if processing finish for the unique Jets feature (either the most separated eyes or the lowest lips). But if a feature, shared with the Sharks face, is processed first, then the decisions can not be made and processing should continue until the termination of the next feature. That complicates the utilization of Jets faces in both the SFT tests (MIC and SIC) and the capacity test. A capacity test for this possibly mixed stopping rule has not been derived yet.

Figure 36: The AND condition, two dimensional face-space defined by the eye-separation and lips-position. The saliency of features is defined by the marginal proximity of each face-feature projection onto the axes with respect to the other group members. Boxed faces belong to Sharks, while faces outside the box are Jets. Jets faces are positioned such that they share one face property with the Sharks. However, in order to recognize Sharks member both face features must be recognized. Jets could be recognized on completion of only one feature that is unique for them. In this example Jets posses one feature that is spread-out the most, while all Sharks faces appear more compact.



In the second part, in addition to the experimental manipulations from the first part another factor was manipulated: whole masked faces. So both whole- and masked-faces are randomly presented. Masked faces are designed using two types of worm-like masks in same way as in the OR condition.

Third part of experiment was a complete identification task identical to the one used in the OR design categorization task where participants have to learn to categorize 4 faces (see Figure 9).

Design and procedure

The identical design was used as in the OR task. In each session participants run 5 blocks of approximately 200 trials each. In contrast to the OR condition, the AND condition demanded a longer learning phase. All participants accomplished a minimum total of 12 sessions during the learning phase of the experiment. However, some participants accomplished a total of 20 sessions (Participants 1 and 4). Because of the duration of whole experimental project, the difficultly of the learning process and participants' availability, different participants accomplished different numbers of learning sessions. However, we monitored their performance during the learning phase and when they reached a learning plateau, on both errors and RTs, we decided when to stop with the learning phase and proceed to the test phase.

Results

Basic Mean RT Analyses

The GLM univariate analysis was conducted on both Sharks and Jets faces separately. However, we will focus our analysis mainly on the Sharks faces, given that they are factorially manipulated and allow for application of the SFT tests (the MIC and SIC tests). The results from GLM analyses for Sharks, for each participant, are presented in the Table 6.

Table 6: GLM univariate analysis was conducted on the Sharks, for different participants (SUB). The main effects and interaction terms are listed in the second column (*Factor*). Degrees of freedom are in the third column (*df*). The *error* row defines degrees of freedom for the F-test error term, for that participant. Each F-test value (*F*) has two degrees of freedom: one from its corresponding row, the other from the error row. The significance level is presented in the column *Sig.*, and the observed power for that effect is in the last column.

	Source	df	F	Sig.	Observed Power		Source	df	F	Sig.	Observed Power
SUB 01	Trial order	1	24.388	.000	.999	**SUB 04**	Trial order	1	86.280	.000	1.000
	Eyes	1	306.153	.000	1.000		Eyes	1	295.647	.000	1.000
	Lips	1	496.686	.000	1.000		Lips	1	474.777	.000	1.000
	Learning Session	9	92.649	.000	1.000		Learning Session	9	40.317	.000	1.000
	Eyes x Lips	1	2.360	.125	.336		Eyes x Lips	1	.457	.499	.104
	Eyes x Learning Session	9	.542	.845	.274		Eyes x Learning Session	9	7.419	.000	1.000
	Lips x Learning Session	9	8.041	.000	1.000		Lips x Learning Session	9	2.801	.003	.963
	Eyes x Lips x Learning Session	9	2.435	.009	.931		Eyes x Lips x Learning Session	9	3.051	.001	.977
	Error	5442					Error	5500			
SUB 02	Trial order	1	2.768	.096	.384	**SUB 05**	Trial order	1	.011	.917	.051
	Eyes	1	208.876	.000	1.000		Eyes	1	45.142	.000	1.000
	Lips	1	600.470	.000	1.000		Lips	1	331.313	.000	1.000
	Learning Session	8	194.411	.000	1.000		Learning Session	6	49.843	.000	1.000
	Eyes x Lips	1	21.270	.000	.996		Eyes x Lips	1	3.466	.063	.461
	Eyes x Learning Session	8	2.943	.003	.957		Eyes x Learning Session	6	.690	.658	.279
	Lips x Learning Session	8	14.463	.000	1.000		Lips x Learning Session	6	2.303	.032	.806
	Eyes x Lips x Learning Session	8	7.836	.000	1.000		Eyes x Lips x Learning Session	6	1.894	.078	.709
	Error	4632					Error	4013			
SUB 03	Trial order	1	.405	.525	.098	**SUB 06**	Trial order	1	3.033	.082	.414
	Eyes	1	133.132	.000	1.000		Eyes	1	215.746	.000	1.000
	Lips	1	660.560	.000	1.000		Lips	1	442.881	.000	1.000
	Learning Session	7	35.012	.000	1.000		Learning Session	7	56.778	.000	1.000
	Eyes x Lips	1	26.029	.000	.999		Eyes x Lips	1	3.818	.051	.497
	Eyes x Learning Session	7	3.873	.000	.984		Eyes x Learning Session	7	1.770	.089	.726
	Lips x Learning Session	7	2.266	.027	.845		Lips x Learning Session	7	8.681	.000	1.000
	Eyes x Lips x Learning Session	7	4.870	.000	.997		Eyes x Lips x Learning Session	7	.470	.857	.209
	Error	4193					Error	4071			

First, we analyzed the overall learning trends for Sharks faces, for all participants: reaction time as a function of the effect of learning session. We found that learning trends are all highly significant, with $p<0.001$ (see Table 6 under the factor Learning Session). The trends of RT as a function of learning session are presented in Figure 37, combined together with the error rate (proportion of errors). All participants exhibited a reduction in the proportion of errors to some asymptotic value as a function of learning sessions. Four participants (1, 2, 3 and 6) exhibited a reduction in both mean RTs and errors. Participant 4 exhibited a trade-off between mean RT and mean proportion of errors for the first three sessions. Participant 4 has a U-shaped mean RT plot as a function of the learning session, in which mean RT increased for the last two sessions, accompanied by a low asymptotic error level.

As it can be seen from Table 6, both manipulated face features, the eye-separation and lip-position factorially combined with feature saliency, exhibited significant main effects, for all participants separately (Table 6 the significance for rows Eyes and Lips, separately). Thus, manipulation of configural face-feature properties produced significant perceptual effects.

Both features also exhibited a significant change in detection over learning sessions for most of the participants (Eyes x Session and Lips x Session) (Table 6).

Of the utmost importance, the MIC test, which was tested by the significance of the Eyes x Lips interaction, was highly significant for Participants 2 and 3; it was marginally significant for Participants 5 and 6, and non-significant for Participants 1 and 4. At this point, we can conclude that 4 of the participants exhibited non-additive effects on the MIC test. Inspection of Figure 38 shows that all MIC contrasts are negative; that is, they are underadditive for these participants. Underadditivity on the MIC level is usually associated with a parallel exhaustive processing architecture, if processes are independent and selectively influenced. Participants 1 and 4 exhibited additivity, which is considered to be, on the MIC level, a signature of serial exhaustive processing.

Figure 37: The AND condition, learning of Sharks faces. Mean RT as a function of learning session, combined together with the error rate (proportion of errors), for all participants. Error bars on the mean RT indicate standard error.

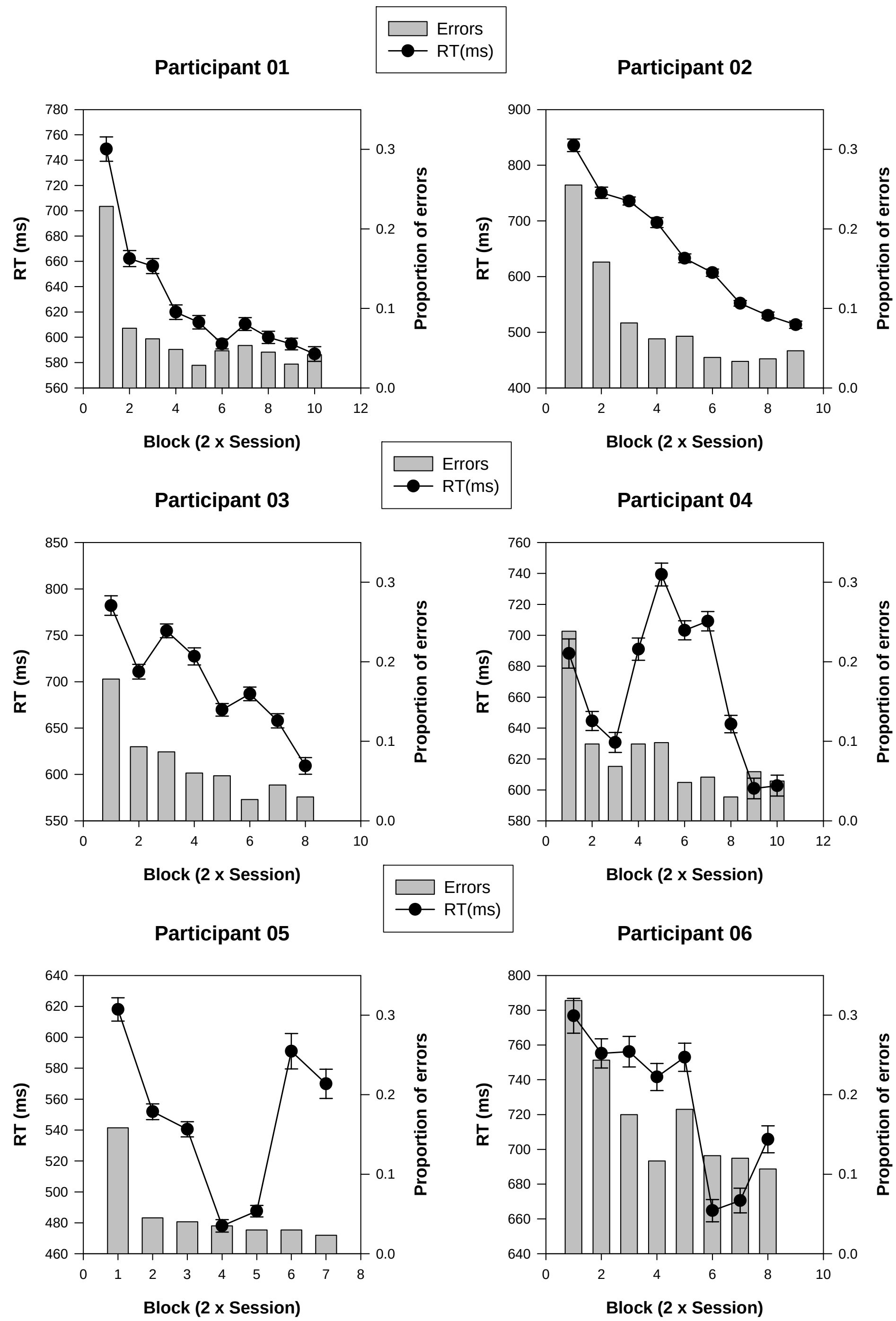
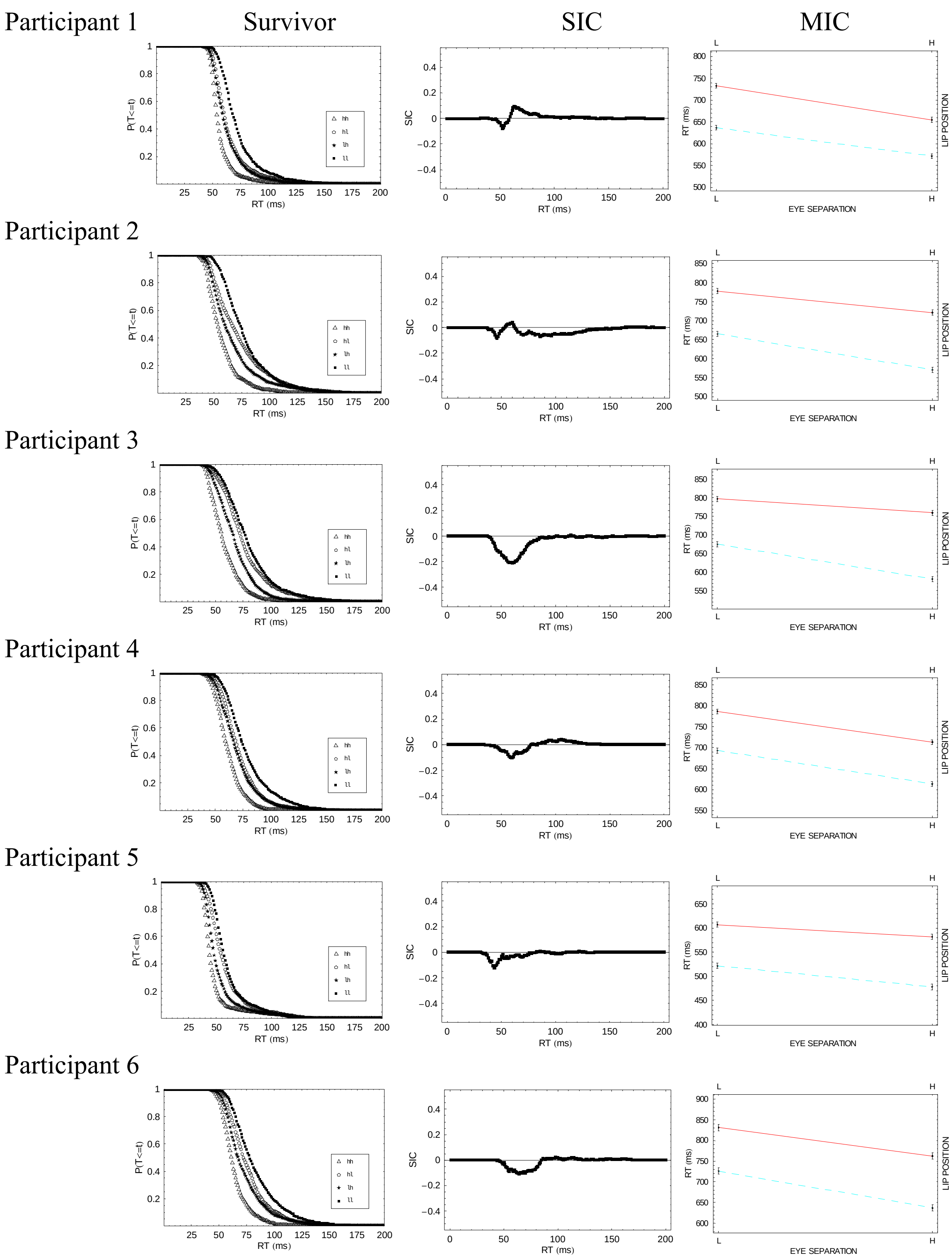


Figure 38: The SFT tests results for the AND condition, for Sharks faces, for all participants. The results are based on all learning sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.



Finally, the 3-way interaction Eyes x Lips x Session, which could indicate a change of MIC over the course of learning, was found to be significant for all participants, with the exception of Participant 5 who reached marginal significance and Participant 6 who was non significant. The participants with a significant 3-way interaction could exhibit a change in architecture during the course of learning.

Learning session and MIC

In order to closely inspect the effect for learning to possible changes in architecture (revealed by changes in the MIC and SIC test scores) we applied similar GLM analysis using the same design but separately for each session for each participant (Table 7). The results supported the overall analysis in Table 6: when broken into sessions, almost all main effects (Eyes, Lips) are significant. Exceptions to this were the main effects on the first learning session for several participants. However, after the first block, all main effects were found to be significant.

Table 7: GLM univariate analysis was conducted on the Sharks, for different participants (SUB) across learning sessions (in successive blocks of rows). The main effects and interaction terms are listed in the second column (*Factor*). Degrees of freedom are in the third column (*df*). The *error* row defines degrees of freedom for F-test error term, for that participant. Each F-test value (*F*) has two degrees of freedom: one from its corresponding row, and the second one from the error row. A significance level is presented in the column *Sig.*, and the observed power for that effect is in the last column.

	Source	df	F	Sig.	Observed Power		Source	df	F	Sig.	Observed Power
Sub 01	Trial order	1	4.060	.044	.520	***Sub 02***	Trial order	1	1.064	.303	.177
Session 01	Eyes	1	12.314	.000	.938	**Session 01**	Eyes	1	4.624	.032	.574
	Lips	1	.269	.604	.081		Lips	1	.844	.359	.150
	Eyes x Lips	1	.163	.687	.069		Eyes x Lips	1	8.482	.004	.828
	Error	462					Error	477			
Session 02	Trial order	1	13.285	.000	.953	**Session 02**	Trial order	1	32.484	.000	1.000
	Eyes	1	32.122	.000	1.000		Eyes	1	4.382	.037	.552
	Lips	1	28.780	.000	1.000		Lips	1	23.648	.000	.998
	Eyes x Lips	1	.439	.508	.101		Eyes x Lips	1	27.773	.000	1.000
	Error	531					Error	522			
Session 03	Trial order	1	9.735	.002	.876	**Session 03**	Trial order	1	.718	.397	.135
	Eyes	1	59.651	.000	1.000		Eyes	1	24.149	.000	.998
	Lips	1	20.014	.000	.994		Lips	1	113.737	.000	1.000
	Eyes x Lips	1	1.844	.175	.273		Eyes x Lips	1	19.110	.000	.992
	Error	548					Error	538			
Session 04	Trial order	1	3.520	.061	.465	**Session 04**	Trial order	1	1.140	.286	.187
	Eyes	1	33.915	.000	1.000		Eyes	1	27.956	.000	1.000
	Lips	1	34.712	.000	1.000		Lips	1	111.397	.000	1.000
	Eyes x Lips	1	4.360	.037	.550		Eyes x Lips	1	13.035	.000	.950
	Error	559					Error	541			
Session 05	Trial order	1	5.068	.025	.613	**Session 05**	Trial order	1	.021	.885	.052
	Eyes	1	33.423	.000	1.000		Eyes	1	40.731	.000	1.000
	Lips	1	81.394	.000	1.000		Lips	1	158.200	.000	1.000
	Eyes x Lips	1	.166	.684	.069		Eyes x Lips	1	1.129	.288	.186
	Error	556					Error	559			
Session 06	Trial order	1	3.510	.062	.464	**Session 06**	Trial order	1	.001	.981	.050
	Eyes	1	55.619	.000	1.000		Eyes	1	28.991	.000	1.000
	Lips	1	96.715	.000	1.000		Lips	1	128.013	.000	1.000
	Eyes x Lips	1	2.411	.121	.341		Eyes x Lips	1	1.445	.230	.224
	Error	554					Error	568			
Session 07	Trial order	1	.333	.564	.089	**Session 07**	Trial order	1	4.312	.038	.545
	Eyes	1	23.655	.000	.998		Eyes	1	56.560	.000	1.000
	Lips	1	106.821	.000	1.000		Lips	1	118.127	.000	1.000
	Eyes x Lips	1	.632	.427	.125		Eyes x Lips	1	5.401	.020	.641
Session 08	Error	551					Error	567			
	Trial order	1	.647	.422	.126	**Session 08**	Trial order	1	.931	.335	.161
	Eyes	1	48.062	.000	1.000		Eyes	1	40.046	.000	1.000
	Lips	1	82.991	.000	1.000		Lips	1	102.739	.000	1.000

	Eyes x Lips	1	10.943	.001	.910		Eyes x Lips	1	3.580	.059	.472
Session 09	Error	562					Error	565			
	Trial order	1	.055	.815	.056	**Session 09**	Trial order	1	2.093	.149	.302
	Eyes	1	32.221	.000	1.000		Eyes	1	19.972	.000	.994
	Lips	1	80.261	.000	1.000		Lips	1	36.699	.000	1.000
	Eyes x Lips	1	1.152	.284	.188		Eyes x Lips	1	.024	.876	.053
Session 10	Error	566					Error	287			
	Trial order	1	.006	.938	.051						
	Eyes	1	34.240	.000	1.000						
	Lips	1	102.632	.000	1.000						
	Eyes x Lips	1	5.136	.024	.619						
	Error	544									

	Source	df	F	Sig.	Observed Power		Source	df	F	Sig.	Observed Power
Sub 03	Trial order	1	1.250	.264	.200	***Sub 04***	Trial order	1	.599	.439	.121
Session 01	Eyes	1	4.477	.035	.560	**Session 01**	Eyes	1	1.760	.185	.263
	Lips	1	80.893	.000	1.000		Lips	1	18.454	.000	.990
	Eyes x Lips	1	12.448	.000	.941		Eyes x Lips	1	.696	.405	.132
	Error	475					Error	506			
Session 02	Trial order	1	.058	.810	.057	**Session 02**	Trial order	1	10.573	.001	.901
	Eyes	1	1.947	.163	.286		Eyes	1	14.207	.000	.964
	Lips	1	74.967	.000	1.000		Lips	1	20.951	.000	.995
	Eyes x Lips	1	33.166	.000	1.000		Eyes x Lips	1	5.086	.025	.615
	Error	562					Error	552			
Session 03	Trial order	1	.011	.918	.051	**Session 03**	Trial order	1	3.118	.078	.422
	Eyes	1	14.585	.000	.968		Eyes	1	24.006	.000	.998
	Lips	1	62.054	.000	1.000		Lips	1	39.265	.000	1.000
	Eyes x Lips	1	10.601	.001	.902		Eyes x Lips	1	9.937	.002	.882
	Error	571					Error	532			
Session 04	Trial order	1	11.963	.001	.932	**Session 04**	Trial order	1	28.790	.000	1.000
	Eyes	1	26.921	.000	.999		Eyes	1	14.331	.000	.966
	Lips	1	118.873	.000	1.000		Lips	1	44.100	.000	1.000
	Eyes x Lips	1	1.378	.241	.216		Eyes x Lips	1	.072	.788	.058
	Error	573					Error	546			
Session 05	Trial order	1	3.711	.055	.485	**Session 05**	Trial order	1	5.992	.015	.686
	Eyes	1	13.872	.000	.961		Eyes	1	22.049	.000	.997
	Lips	1	95.864	.000	1.000		Lips	1	62.762	.000	1.000
	Eyes x Lips	1	.419	.518	.099		Eyes x Lips	1	.876	.350	.154

	Error	572					Error	568			
Session 06	Trial order	1	.037	.847	.054	**Session 06**	Trial order	1	.206	.650	.074
	Eyes	1	45.689	.000	1.000		Eyes	1	23.663	.000	.998
	Lips	1	123.430	.000	1.000		Lips	1	62.084	.000	1.000
	Eyes x Lips	1	.852	.356	.151		Eyes x Lips	1	.567	.452	.117
	Error	569					Error	549			
Session 07	Trial order	1	41.249	.000	1.000	**Session 07**	Trial order	1	25.851	.000	.999
	Eyes	1	25.022	.000	.999		Eyes	1	57.094	.000	1.000
	Lips	1	95.991	.000	1.000		Lips	1	38.235	.000	1.000
	Eyes x Lips	1	3.375	.067	.450		Eyes x Lips	1	2.051	.153	.298
	Error	581					Error	571			
Session 08	Trial order	1	.988	.321	.168	**Session 08**	Trial order	1	24.770	.000	.999
	Eyes	1	20.338	.000	.994		Eyes	1	29.968	.000	1.000
	Lips	1	56.835	.000	1.000		Lips	1	107.419	.000	1.000
	Eyes x Lips	1	1.078	.300	.179		Eyes x Lips	1	2.519	.113	.354
	Error	283					Error	557			
						Session 06	Trial order	1	33.573	.000	1.000
							Eyes	1	81.565	.000	1.000
							Lips	1	64.714	.000	1.000
							Eyes x Lips	1	.479	.489	.106
							Error	545			
						Session 09	Trial order	1	.123	.725	.064
							Eyes	1	89.258	.000	1.000
							Lips	1	64.895	.000	1.000
							Eyes x Lips	1	5.399	.021	.640
							Error	565			

	Source	df	F	Sig.	Observed Power		Source	df	F	Sig.	Observed Power
Sub 05	Trial order	1	30.118	.000	1.000	***Sub 06***	Trial order	1	.953	.329	.164
Session 01	Eyes	1	2.961	.086	.404	**Session 01**	Eyes	1	30.720	.000	1.000
	Lips	1	21.958	.000	.997		Lips	1	1.060	.304	.177
	Eyes x Lips	1	1.688	.194	.254		Eyes x Lips	1	.354	.552	.091
	Error	544					Error	462			
Session 02	Trial order	1	5.474	.020	.646	**Session 02**	Trial order	1	1.677	.196	.253
	Eyes	1	7.906	.005	.802		Eyes	1	24.559	.000	.999
	Lips	1	151.416	.000	1.000		Lips	1	38.124	.000	1.000
	Eyes x Lips	1	17.194	.000	.985		Eyes x Lips	1	1.105	.294	.182
	Error	569					Error	499			

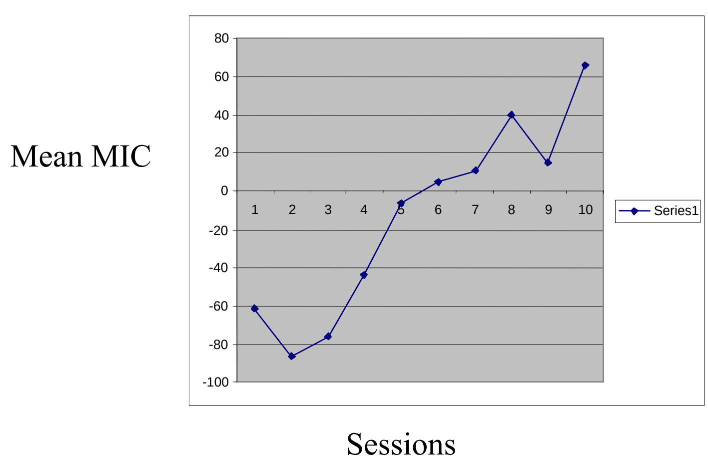
Session 03	Trial order	1	51.931	.000	1.000	**Session 03**	Trial order	1	11.828	.001	.930
	Eyes	1	6.206	.013	.701		Eyes	1	41.398	.000	1.000
	Lips	1	142.670	.000	1.000		Lips	1	68.448	.000	1.000
	Eyes x Lips	1	5.223	.023	.626		Eyes x Lips	1	3.339	.068	.446
	Error	579					Error	495			
Session 04	Trial order	1	.682	.409	.131	**Session 04**	Trial order	1	.696	.404	.132
	Eyes	1	21.845	.000	.997		Eyes	1	26.287	.000	.999
	Lips	1	135.185	.000	1.000		Lips	1	74.629	.000	1.000
	Eyes x Lips	1	.506	.477	.110		Eyes x Lips	1	2.095	.148	.304
	Error	578					Error	511			
Session 05	Trial order	1	.249	.618	.079	**Session 05**	Trial order	1	.077	.782	.059
	Eyes	1	48.258	.000	1.000		Eyes	1	28.908	.000	1.000
	Lips	1	128.385	.000	1.000		Lips	1	50.467	.000	1.000
	Eyes x Lips	1	.164	.686	.069		Eyes x Lips	1	.531	.467	.112
	Error	579					Error	519			
Session 06	Trial order	1	169.481	.000	1.000	**Session 06**	Trial order	1	5.047	.025	.611
	Eyes	1	3.288	.070	.441		Eyes	1	39.067	.000	1.000
	Lips	1	13.414	.000	.955		Lips	1	61.838	.000	1.000
	Eyes x Lips	1	.142	.706	.066		Eyes x Lips	1	.112	.739	.063
	Error	578					Error	528			
Session 07	Trial order	1	33.444	.000	1.000	**Session 07**	Trial order	1	.351	.554	.091
	Eyes	1	4.893	.027	.598		Eyes	1	12.359	.000	.939
	Lips	1	39.113	.000	1.000		Lips	1	91.166	.000	1.000
	Eyes x Lips	1	.775	.379	.142		Eyes x Lips	1	.309	.579	.086
	Error	580					Error	528			
						Session 08	Trial order	1	8.420	.004	.825
							Eyes	1	35.274	.000	1.000
							Lips	1	137.927	.000	1.000
							Eyes x Lips	1	.006	.936	.051
							Error	522			

The significance of the interaction between the two (Eyes x Lips) exhibited variability over the learning sessions. However, this apparent variability of significance of the interaction over different session was found to be associated with a lawful change of the MIC over the learning sessions. In fact, it can be observed from Table 7, that the MIC scores tended to increase from some negative value toward zero, and sometimes it reached a positive value, as a function of learning session. So, we investigated the relationship between learning session and the change of MIC.

We previously showed a significant three-way interaction (Eyes x Lips x Sessions) for most participants which directly indicates that the MIC values should exhibit some change as a function of learning session. In order to test this, we ran a multiple regression analysis using subjects' MIC scores as the dependent variable, while the session, the mean RT, and their interaction were independent variables. The proportion of explained variability of the MIC change was R^2=0.58, ($F(2,35)$=22.83), with both variables, the session and mean RT, contributing significantly. So, almost 60% of the individual variability of MIC values can be explained by the amount of learning and mean reaction time across learning sessions.

When we average over different participants, these predictors can explain 94% of the MIC variability, due to function smoothing after averaging over participants (see Figure 39) (R^2=.94, $F(2,7)$=50.172, $p<0.001$); both predictor variables (learning session and mean RT for each session) entered the regression analysis and were significant at the $p<0.05$ level.

Figure 39: MIC value as a function of the learning session, averaged over participants.

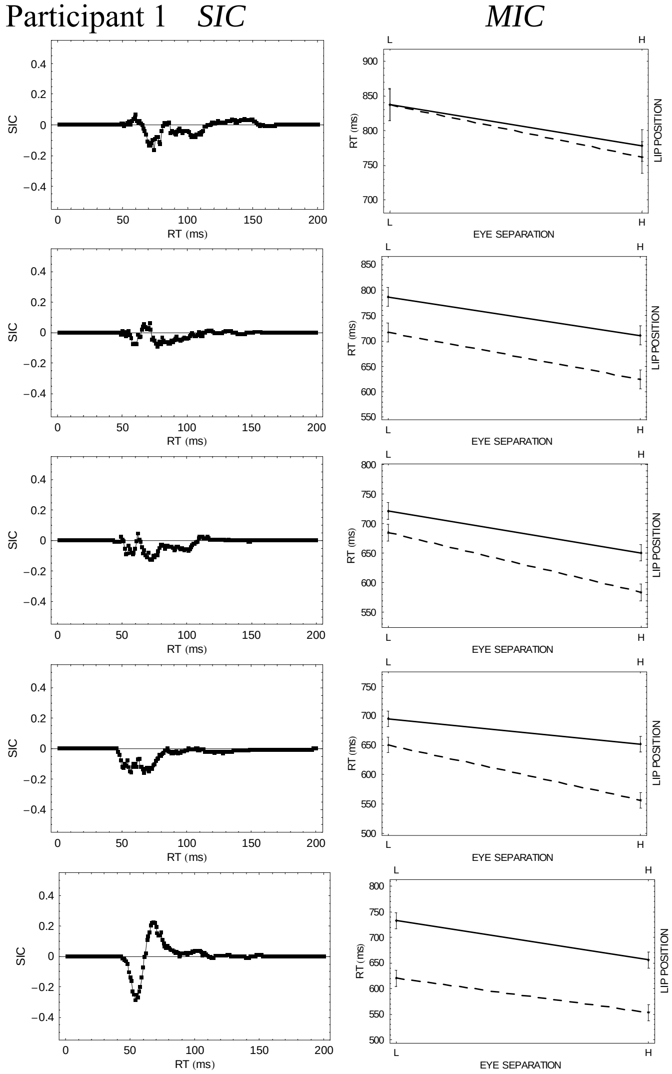


Mean and Survivor Interaction Contrast Functions

In figure 40 we show the MIC results for different participants, combined with their corresponding SIC functions. Note that both tests are calculated over all learning session. We also provide additional corresponding figures of the survivor functions used to calculate the SIC function (Figure 40 on the left).

Figure 40: The SFT tests results for the AND condition, for Shark faces, for all participants. The results are broken down for each participant across the learning sessions. The survivor interaction contrast functions (SIC) are in the left column. The MIC values are presented in the right column; dotted lines connect points with high lips-position saliency level (H). The learning sessions are presented in rows

Participant 1 *SIC* *MIC*



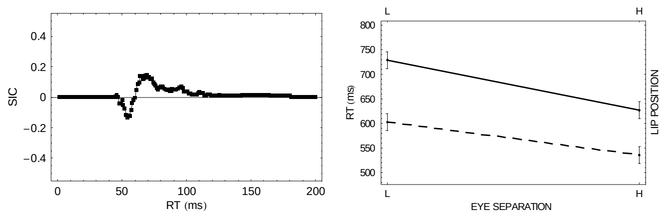
Session 1

Session 2

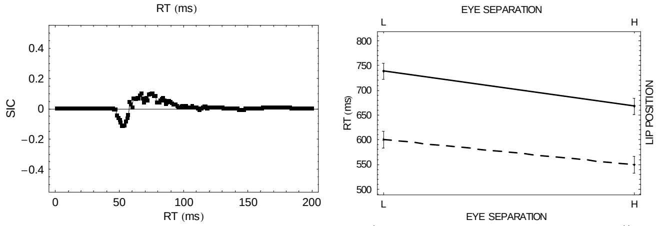
Session 3

Session 4

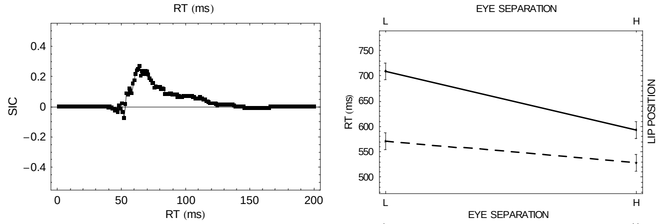
Session 5



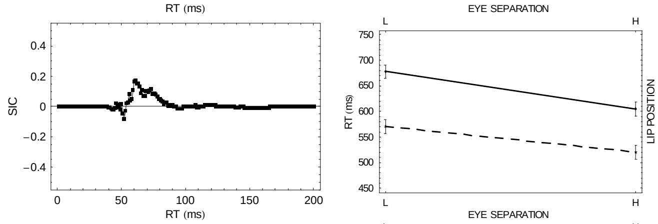
Session 6



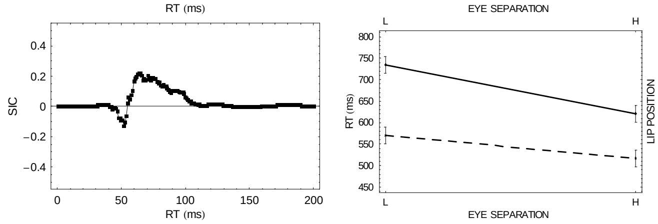
Session 7



Session 8

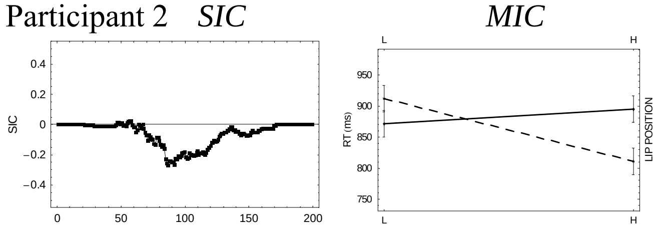


Session 9

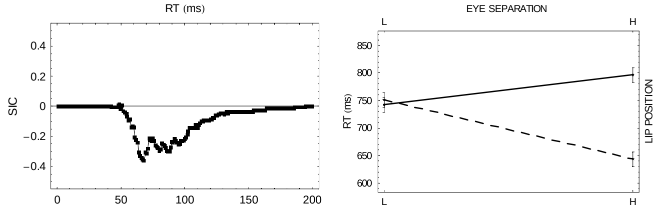


Session 10

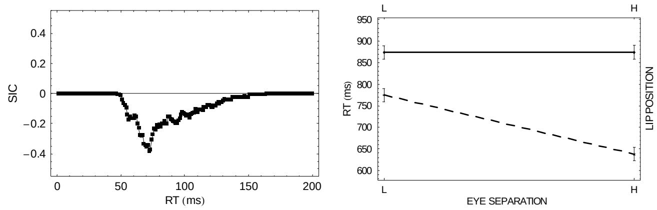
Participant 2 *SIC* *MIC*



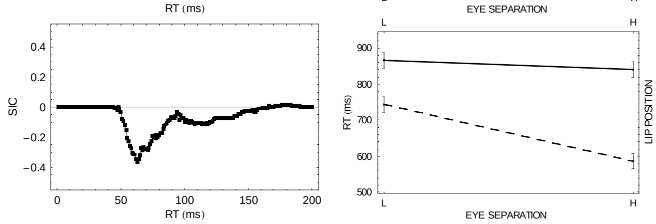
Session 1



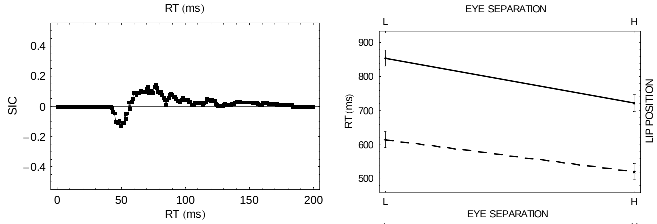
Session 2



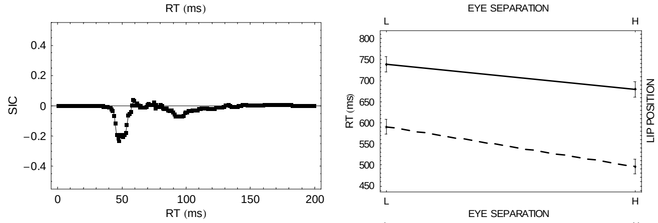
Session 3



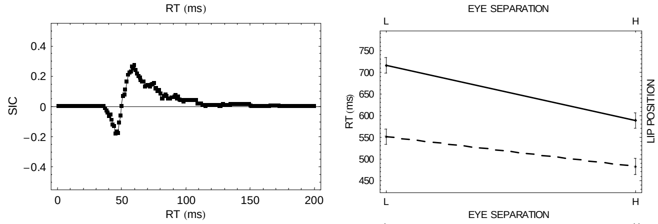
Session 4



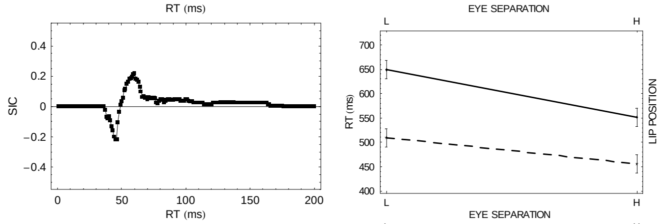
Session 5



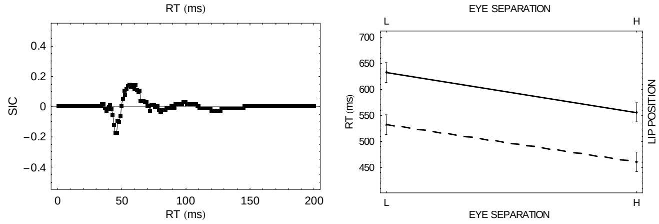
Session 6



Session 7

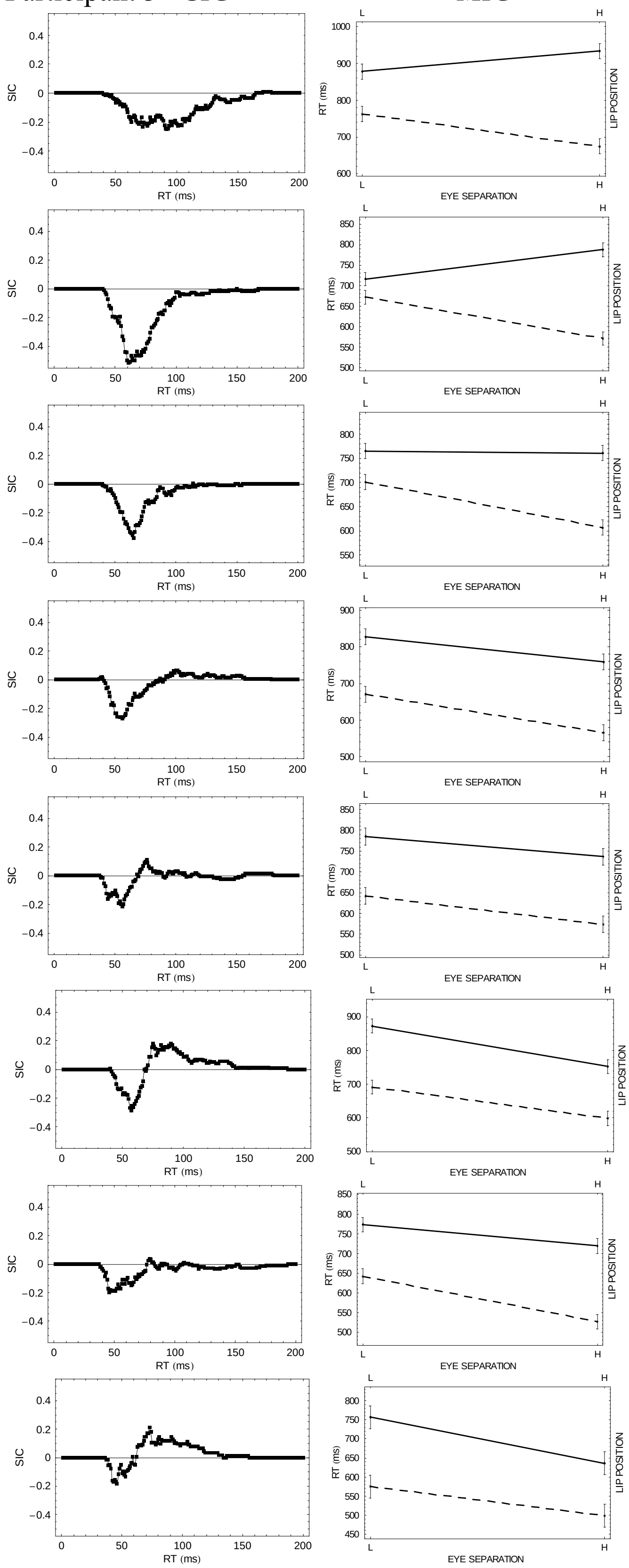


Session 8



Session 9

Participant 3 *SIC* *MIC*



Session 1

Session 2

Session 3

Session 4

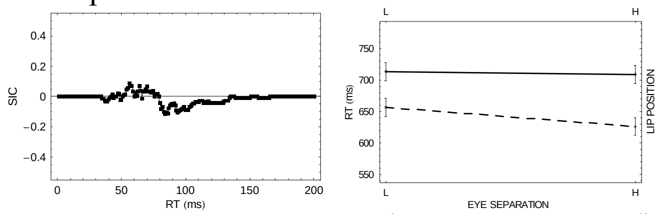
Session 5

Session 6

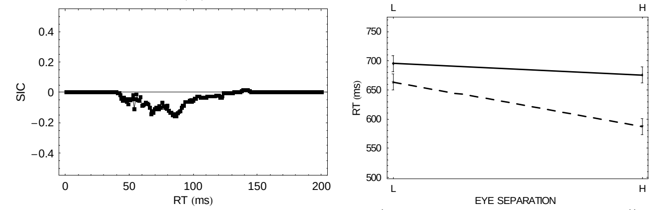
Session 7

Session 8

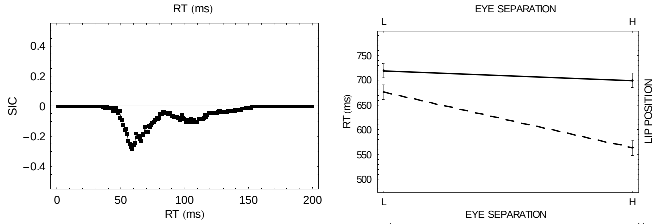
Participant 4 *SIC* *MIC*



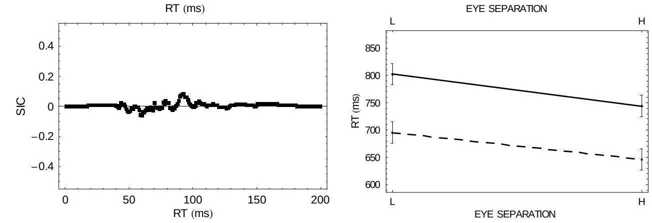
Session 1



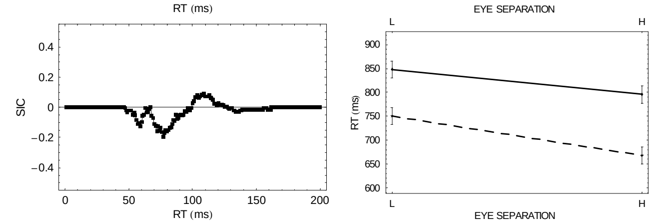
Session 2



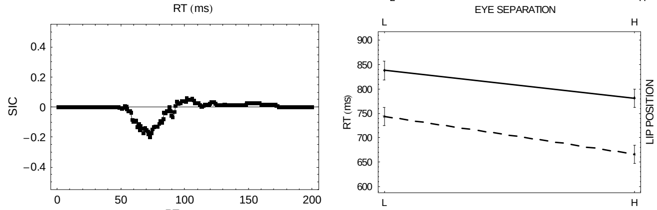
Session 3



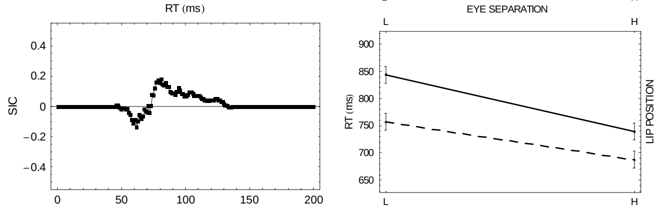
Session 4



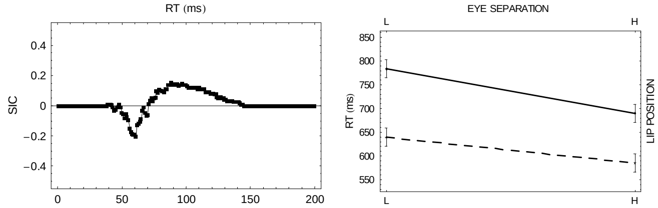
Session 5



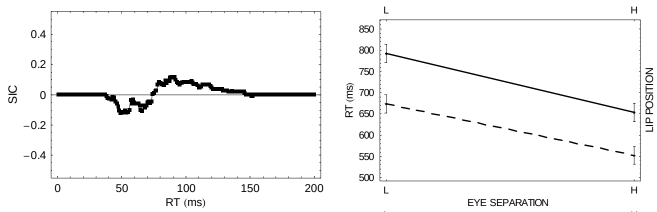
Session 6



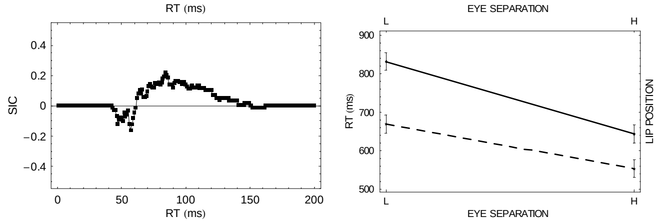
Session 7



Session 8

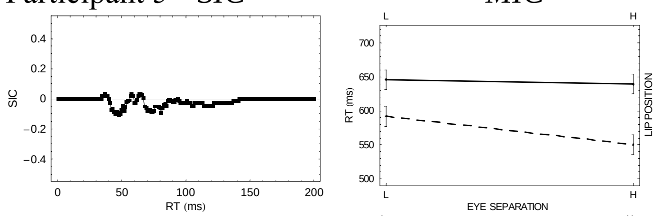


Session 9

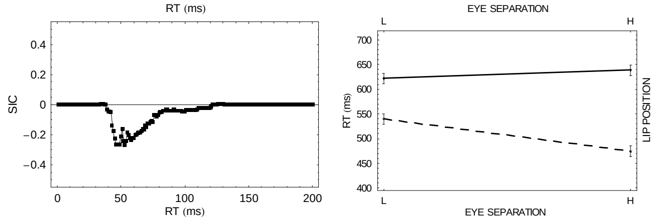


Session 10

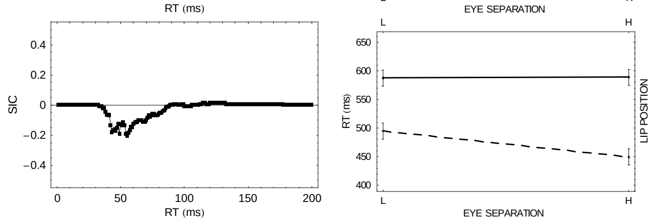
Participant 5 *SIC* *MIC*



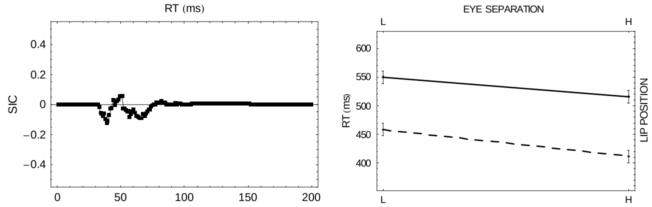
Session 1



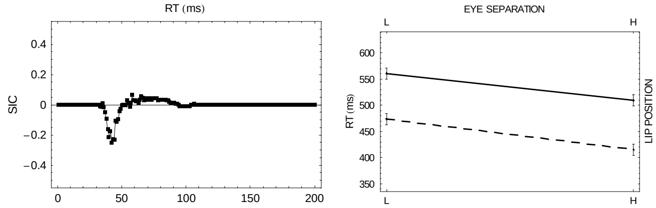
Session 2



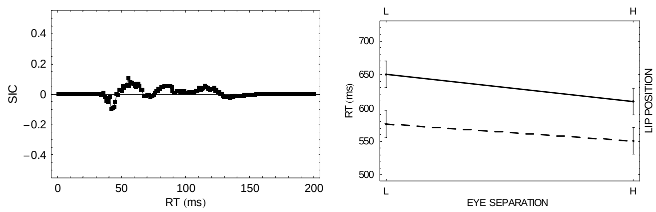
Session 3



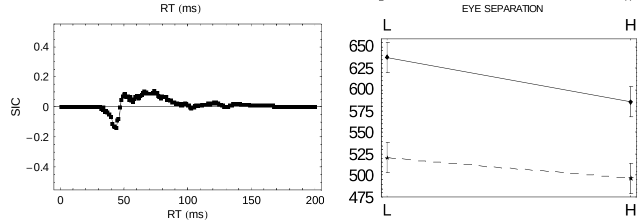
Session 4



Session 5

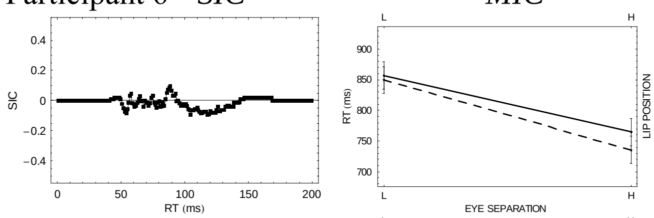


Session 6

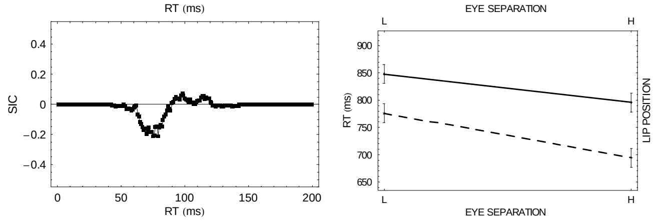


Session 7

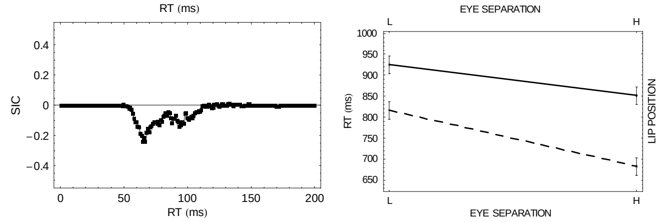
Participant 6 *SIC* *MIC*



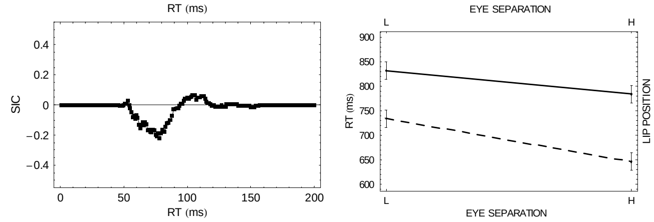
Session 1



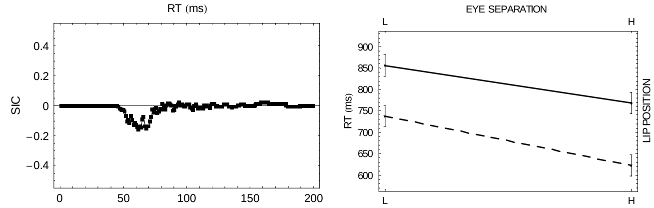
Session 2



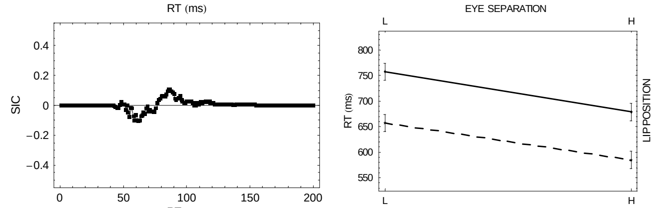
Session 3



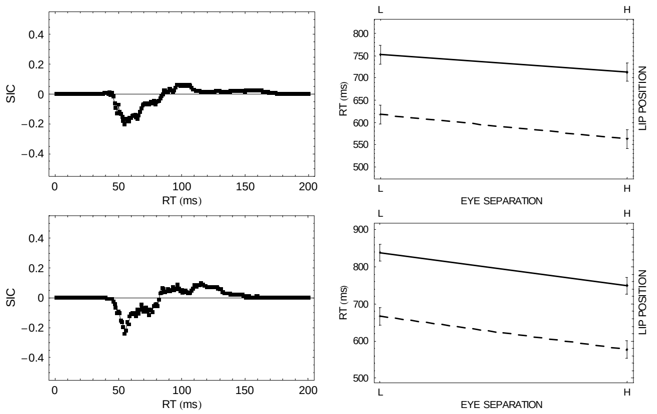
Session 4



Session 5



Session 6



Session 7

Session 8

From the shapes of the SIC functions, we see that the conclusions from this test support the basic MIC test using the GLM analyses (see above). Participants 1 and 4 exhibited S-shaped SIC functions, which corresponds to the serial exhaustive architecture, on stronger level than the ordering of means. Other participants' SIC functions are mainly negative, which is strong indication of parallel exhaustive processing. Participant 2 exhibited some mixture of both functions, where the early part of the SIC is S-shaped and the later part is predominantly negative. The shape of the SIC function warns us about possible scaling issues when combining several SIC functions which are calculated from different sessions, that could be affected by different levels of learning. So, in the next session we decided to break down the overall SIC function into several functions for each learning session.

In Figure 40 we broke down the both the MIC and SIC into different learning sessions. In agreement with previous analyses based on GLM tests, and following overall finding of underadditivity on both the MIC and SIC levels, we revealed the same signature of parallel maximum time (exhaustive) processing architecture as in earlier sessions, for different participants. Again, the shapes of the SIC functions paralleled the findings on the MIC level, where we discovered lawful changes of MIC over learning sessions. In fact, for most participants, there is a common scenario that develops over the sessions. In the first learning sessions, participants usually show unordered means (such that LL>LH≈HL>HH does not hold). We suggest that the initial learning set the processes in an appropriate frame of constraints and pushed toward optimal performance. The process of learning established the desired ordering of means and survivor functions in later sessions. Around the third session a negative SIC function is clearly established, suggesting that parallel exhaustive architecture was engaged, for almost all participants. Then, later in learning there is a clear tendency for reduction in the MIC value to zero which reaches a positive value for the later learning sessions, for most of the participants. This lawful change in the MIC value was accompanied by change of the shape of the SIC function, which made a transition from mainly negative to a positive, S-shaped function. And in many cases, processing did produce MIC overadditivity, which is exhibited as a positive S-shaped SIC function (in Table 7 we can see that some of the MIC tests with positive values are significant).

Difference between blocked and mixed whole-face conditions

As the null hypothesis, we assumed that there was no difference on the mean level between blocked and mixed conditions when whole faces are used. Again, we tested the relationship between the utilization of part-based faces and whole-face information.

In order to test this hypothesis we ran paired sample t-tests on the mean difference between processing whole faces in the first part (the SFT test) of the learning session and processing time for whole faces in the mixed condition (from the capacity test part) for each participants. We found that whole faces are processed faster in the blocked than in the mixed conditions ($t(5)=-6.468$, $RT_{blocked}$=667ms, SD =66ms, and RT_{mixed}=746ms, SD = 82ms), and that this finding is consistent over all participants. We conclude that the processing of whole faces in the two conditions is altered by the context of part-based (masked) information. The major consequence of this finding is on calculation of the capacity functions in the next section. Since there is a significant difference of processing of whole faces depending on the presence of a part-face context in the experiment, we will utilize calculation of two different capacity coefficient functions, each of which uses a different whole-face calculation (from either the blocked or mixed conditions).

Capacity Coefficient Functions (CCFs)

We calculated 4 different capacity coefficient functions for each participant, for each stimulus category. In Figures 41, 42, 43 and 44 we present the calculated CCFs, along with bootstrapped 90% percentile confidence intervals, for both Sharks faces (pooled together into one condition). So for each participant, we calculated 8 CCFs, for each learning session. Confidence intervals delimit the area in which the capacity coefficient function could fall when the same experiment is repeated.

We assume that in the case of super capacity ($C(t)>1$), both confidence interval bounds should swing above the value of one. In the case of unlimited capacity, the CCF value of one should be in between the two confidence interval bounds; in other words, one bound is above $C(t)=1$ and the other one is below $C(t)=1$. And in the case of limited capacity, both confidence interval bounds should fall below $C(t)=1$.

It is also possible that at different points in the time scale, the generated confidence intervals are some combination of cases, such that at some point in time they are both above one, and for a later time they could contain C(t)=1.

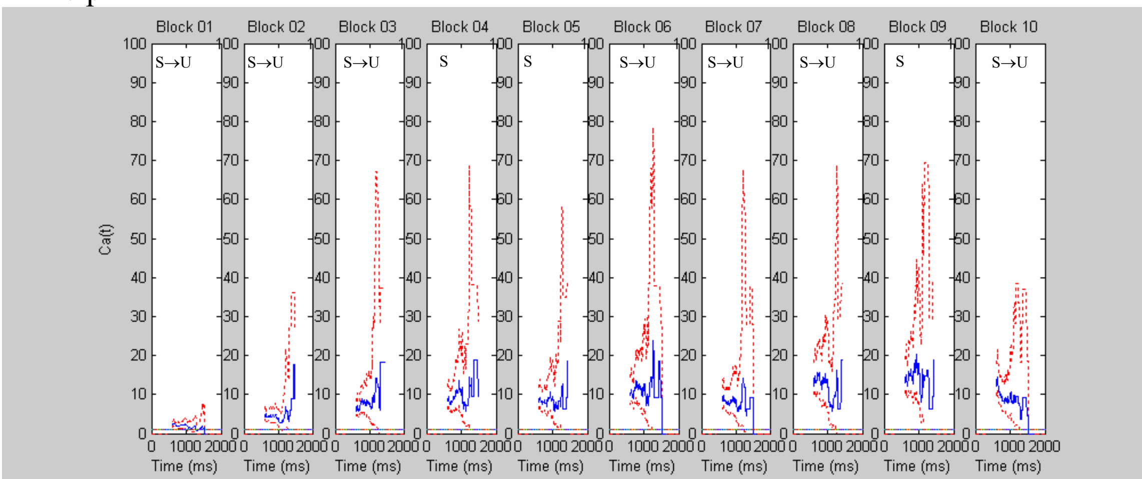
In Figures 41-44, we plot CCFs (4) for each participant (6), for Sharks only. Remember that Sharks were the group of faces that permitted application of the SFT test. Jets did not permit SFT analysis: in order to decide that a face was a Jet, a combination of OR and AND processing could be used. If participant processed a unique Jet's feature first, then he/she can terminate, but if a shared (with Sharks) face feature is processed first, then the second feature must be processed as well.

We also used scales on the y-axis that are common for all figures in the experiment, so a reader can see changes of magnitude as a function of learning session. On each figure, we denoted by a blue bold line the calculated CCF, and with red dotted lines we present the bootstrapped 90% confidence intervals. $C_a(t)=1$ reference value is denoted by a horizontal line. Note that violation of C(t)=1 in any direction (super capacity or limited capacity) must show at least at some point a violation of both bootstrapped confidence interval bounds. That is, they both must swing above or below C(t)=1.

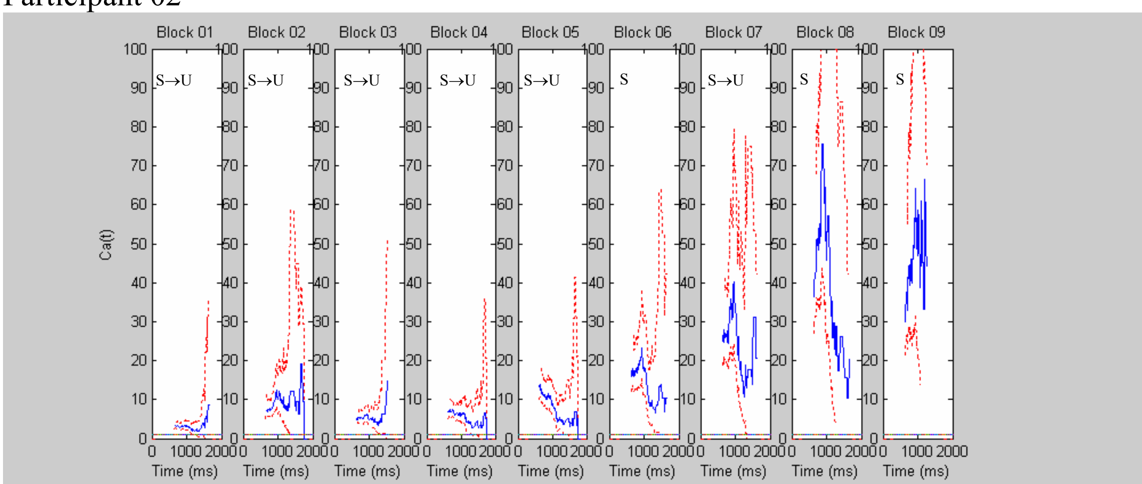
In some figures, it could be difficult to verify by eye violations of C(t)=1 for the confidence intervals. Although that could be solved by rescaling of the y-axis, the reader will miss information about magnitude over the learning sessions. Again, we scale the y-axis such that all CCFs for one logical conditions (OR or AND) are comparable, given the largest observed CCF. In order to increase the understanding of Figures 41-44, we added the statistical conclusions concerning violations in each small figure: if the CCF was super, unlimited or limited capacity then we used uppercase letters S, U and L, respectively. If a transition was observed, that is for some time a CCF could be super capacity then unlimited capacity for rest of the time, we used an arrow symbol. So, the notation S→U→L means that the CCF exhibited all capacity states over the course of time. In some rare cases, the order could be more complex, but on the whole this is what we can observe in this study. It was previously demonstrated that the $C_a(t)$ function for AND processing usually shows an increasing trend as a function of time (Townsend & Wenger, 2004). Note however that this is true in most cases in the observed CCF only, but after bootstrapping the lower confidence interval bound showed a final decreasing trend as a function of time.

Figure 41: The **absolute learning whole-blocked CCFs**, for the **Sharks members**, across the learning sessions (Block), separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of the C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

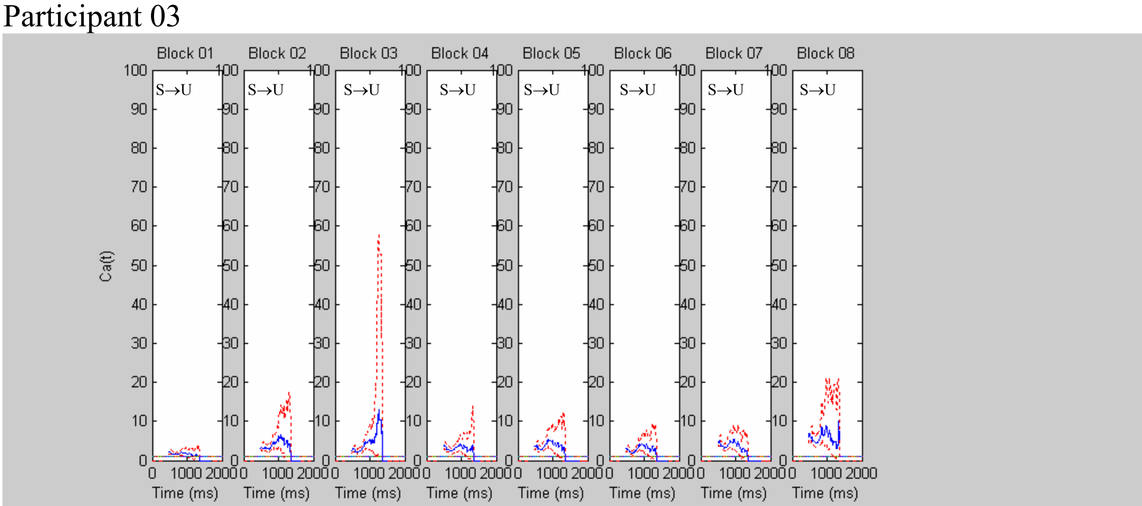
Participant 01



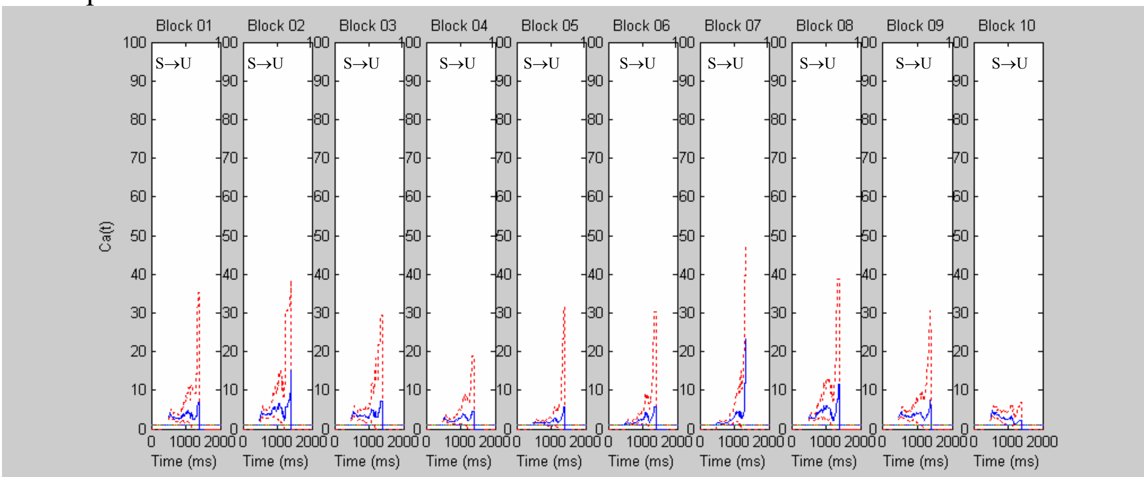
Participant 02



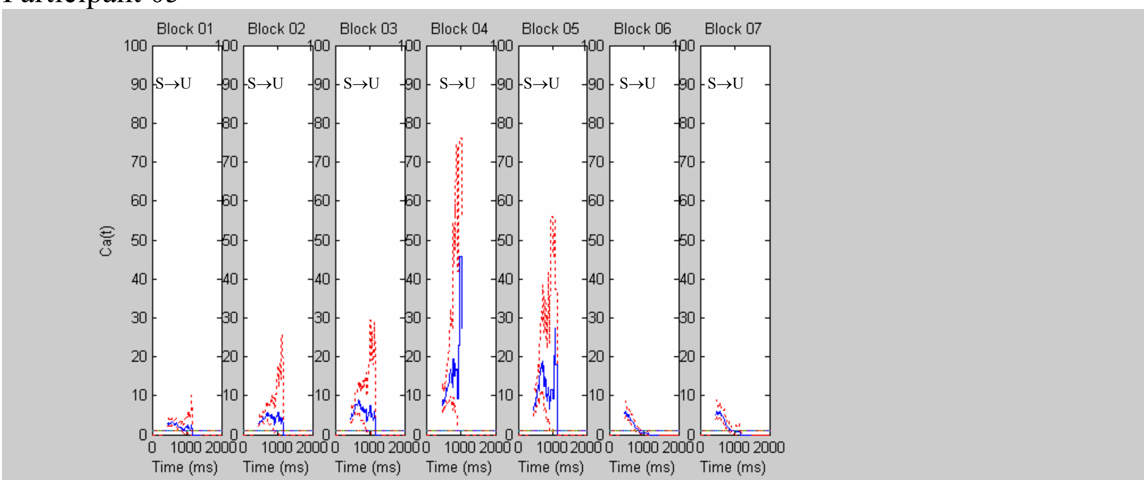
Participant 03



Participant 04



Participant 05



Participant 06

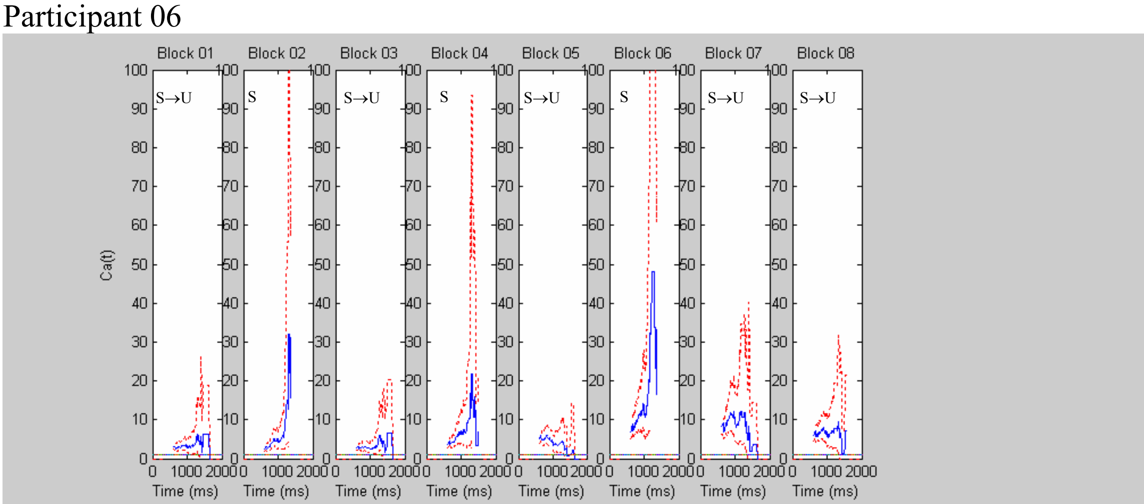
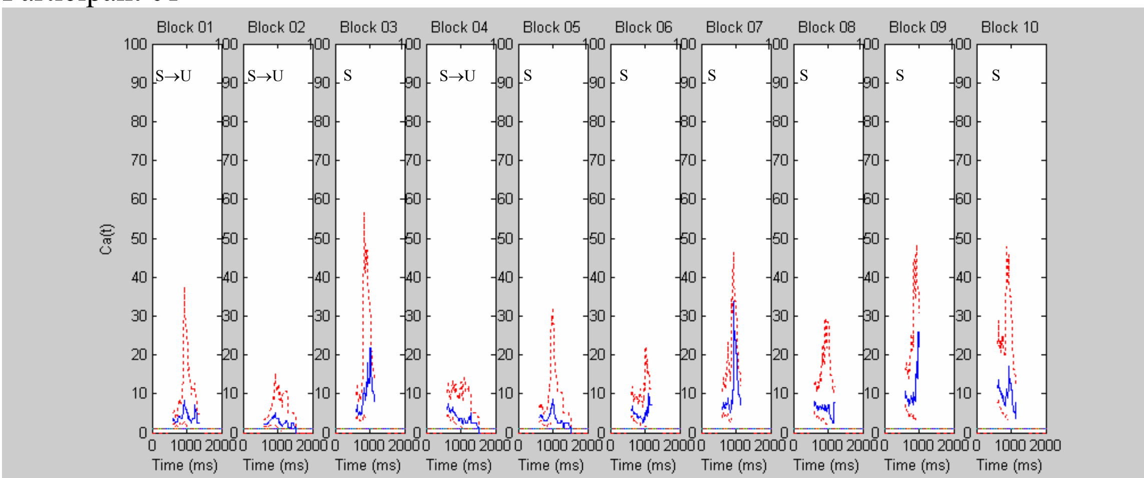
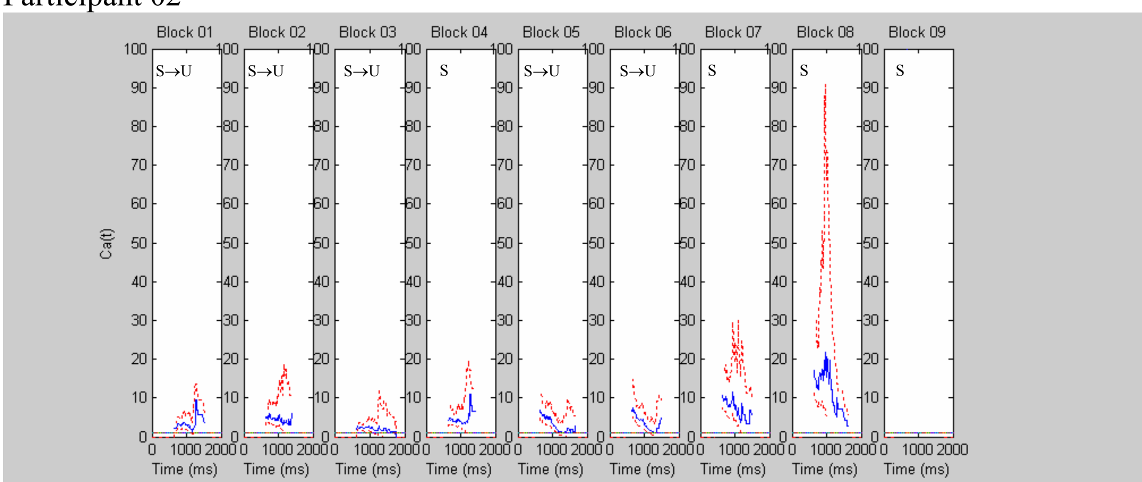


Figure 42: The **Absolute learning whole-mixed CCFs**, for the **Sharks members**, across the learning sessions (Block), separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of the C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

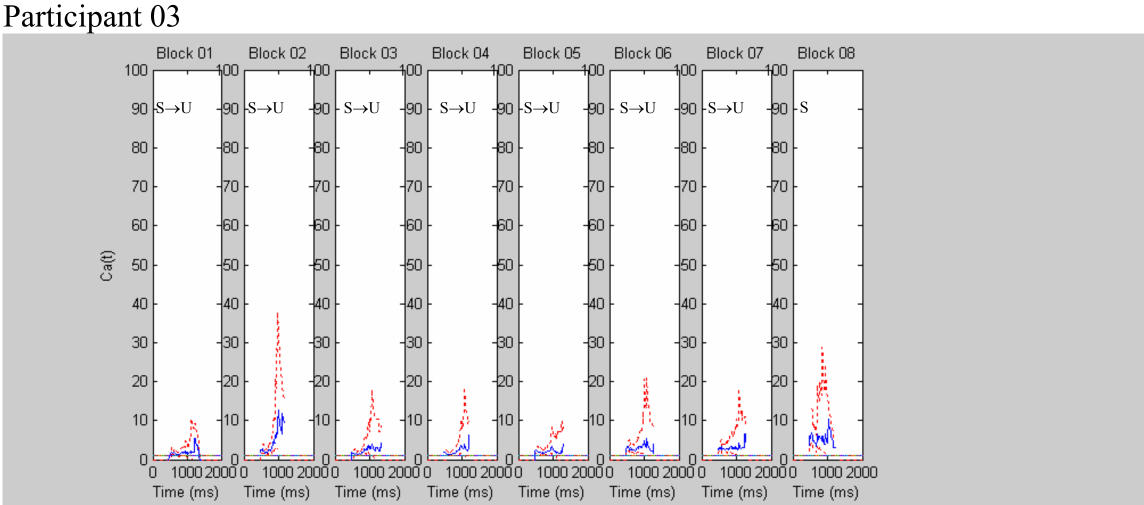
Participant 01



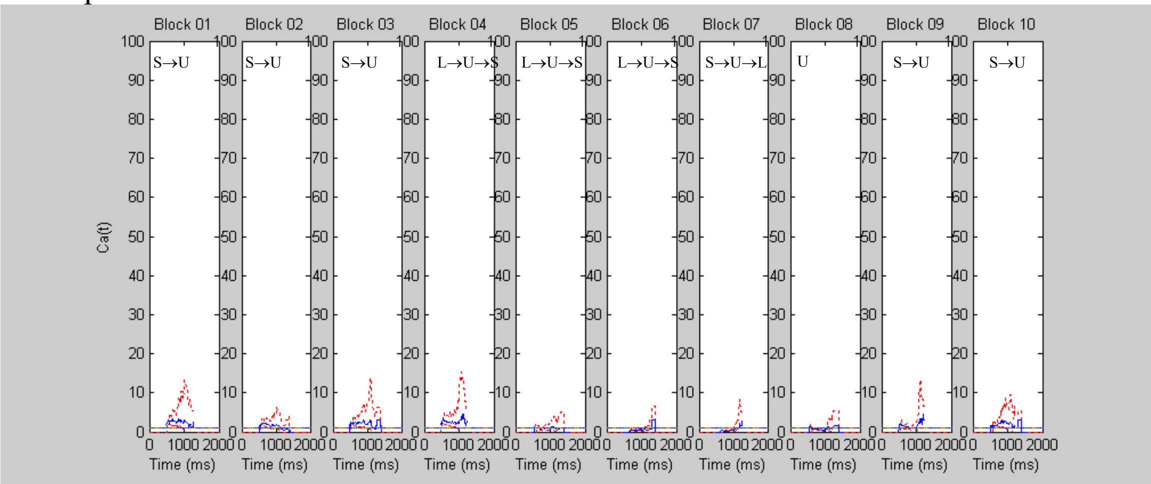
Participant 02



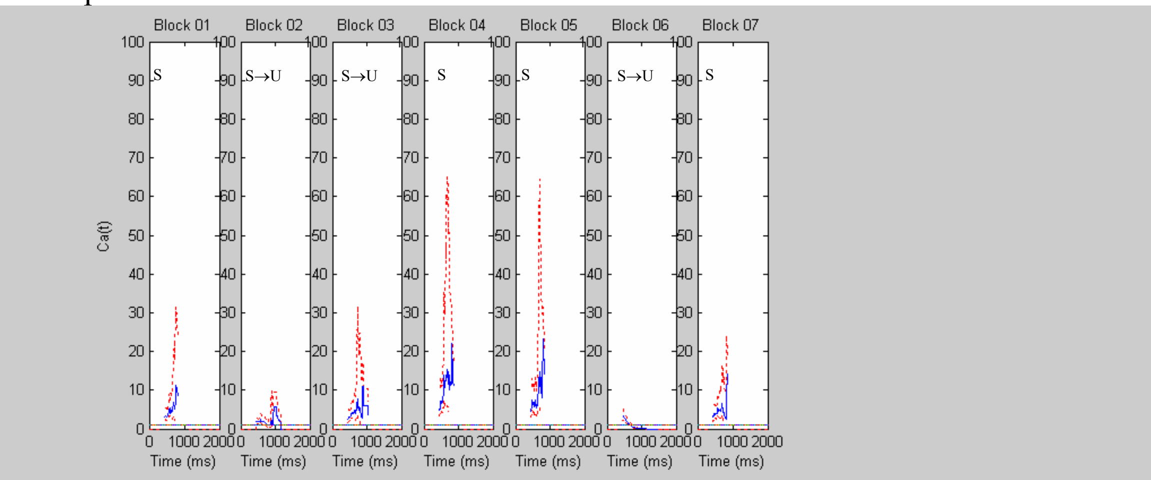
Participant 03



Participant 04



Participant 05



Participant 06

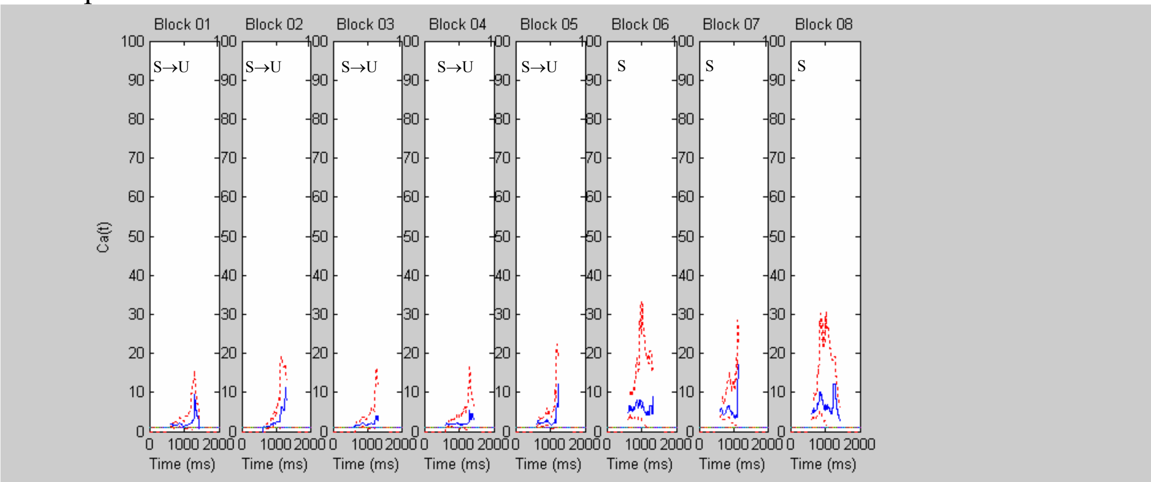
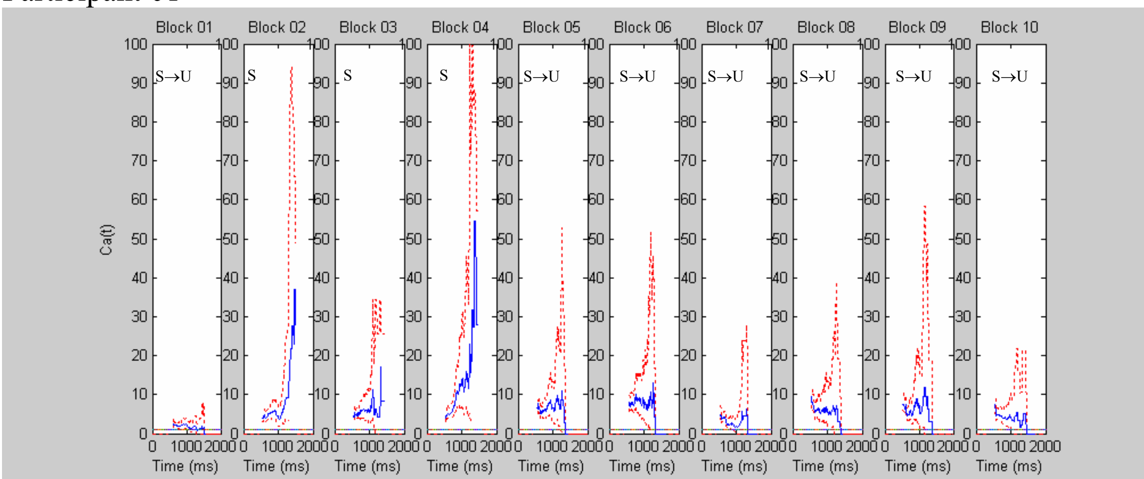
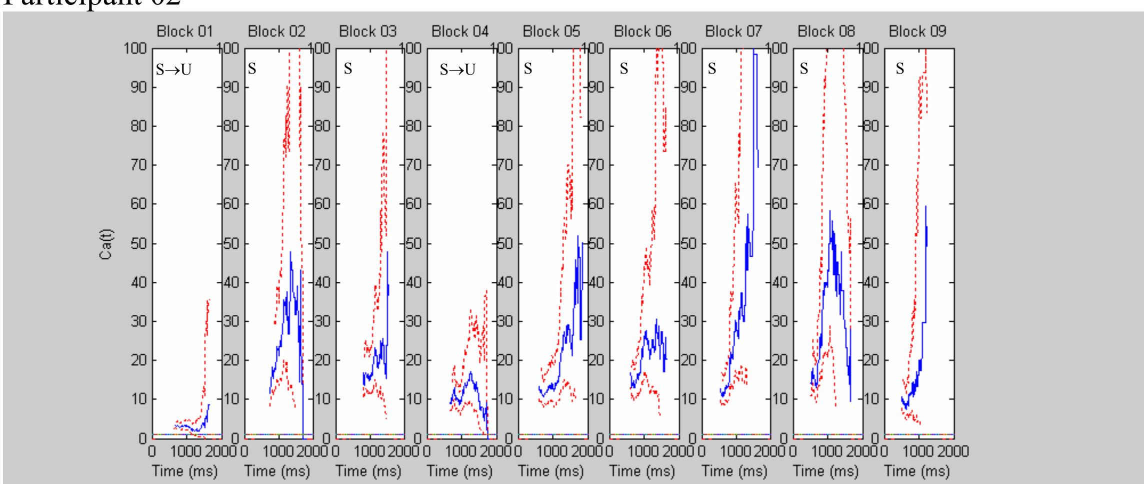


Figure 43: The **Relative learning whole-blocked CCFs**, for the **Sharks members**, across the learning sessions (Block), separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of the C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

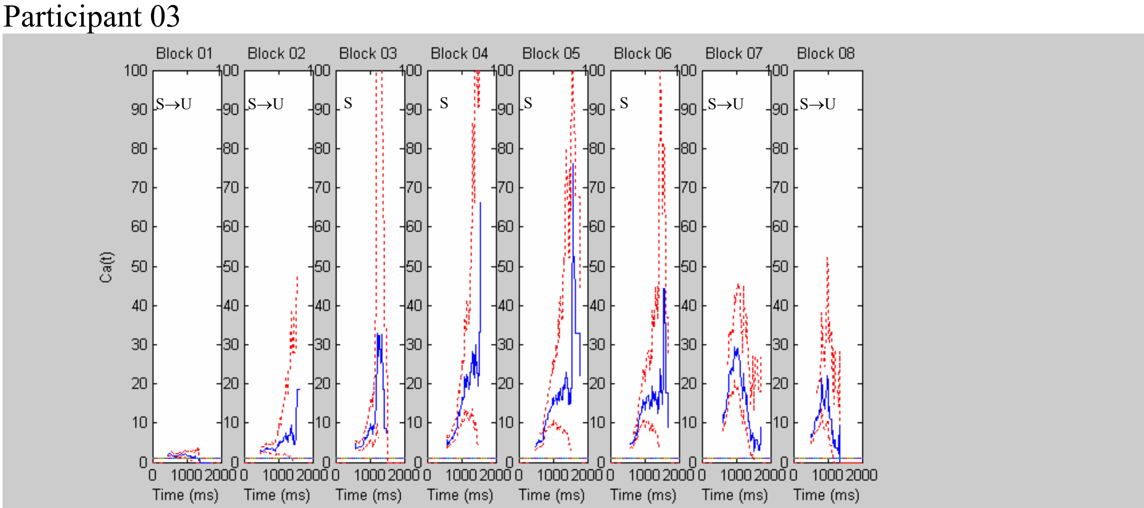
Participant 01



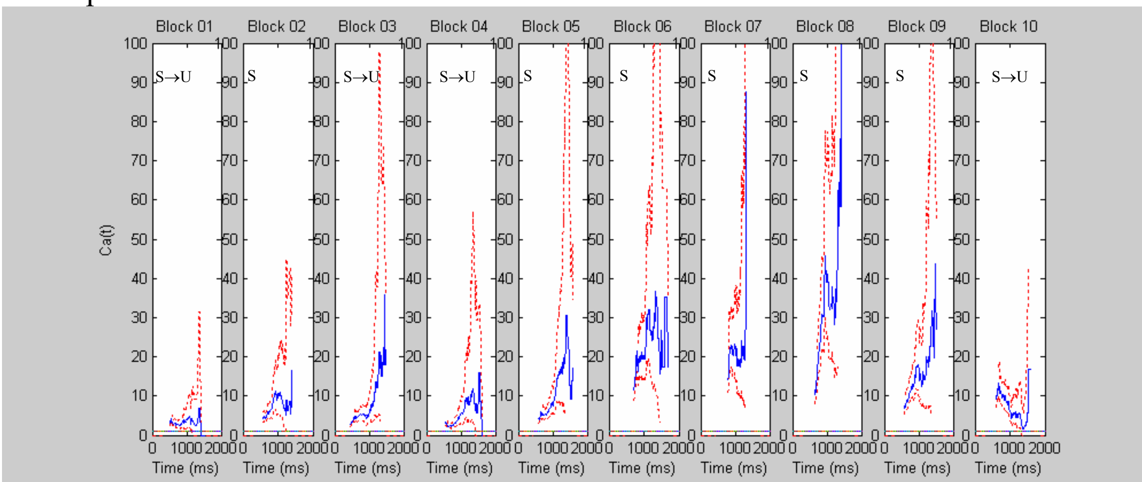
Participant 02



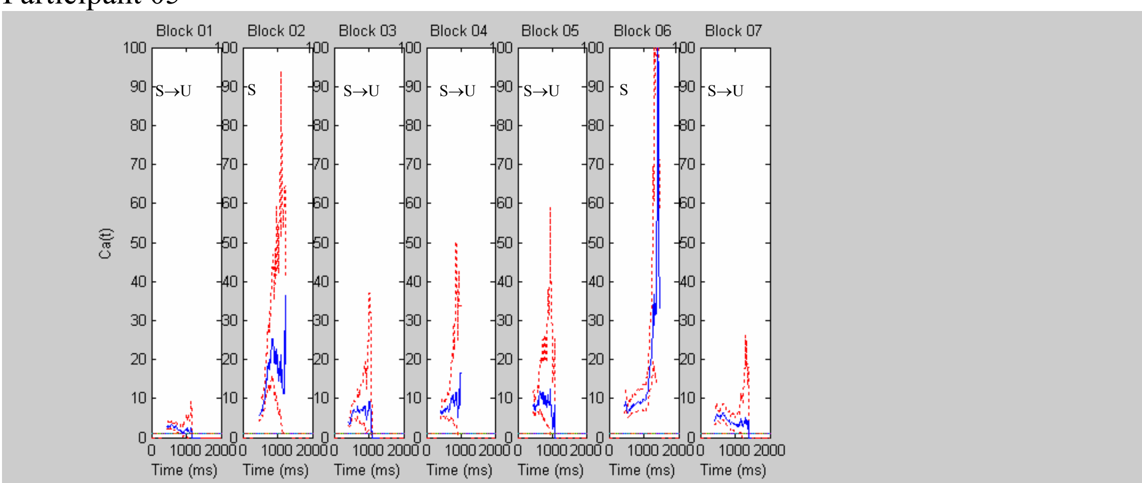
Participant 03



Participant 04



Participant 05



Participant 06

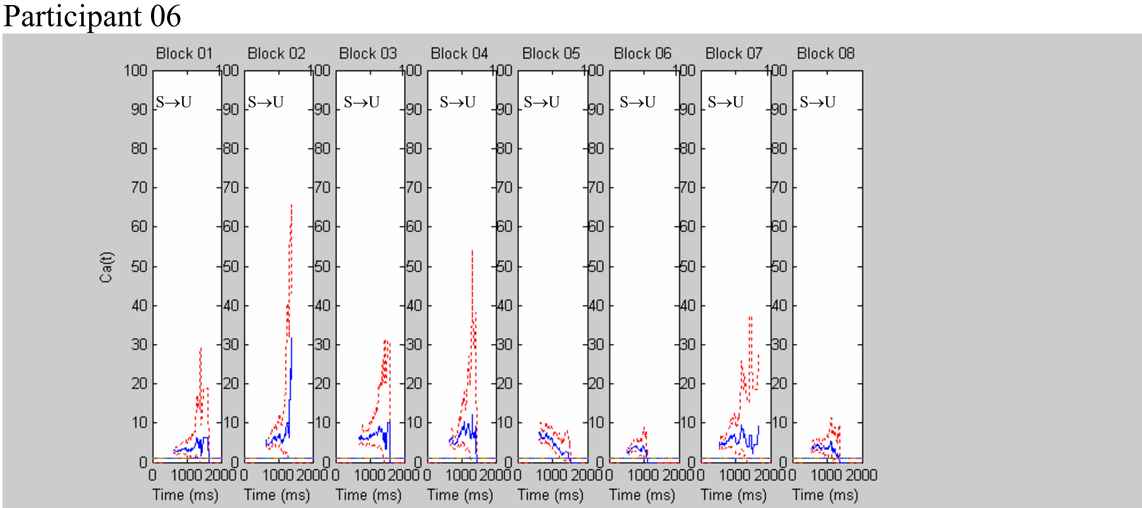
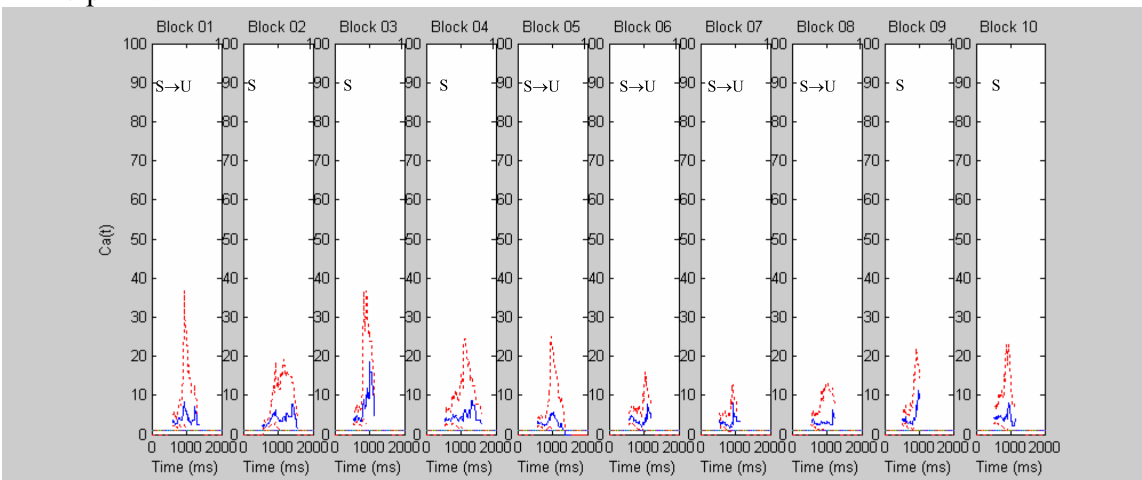
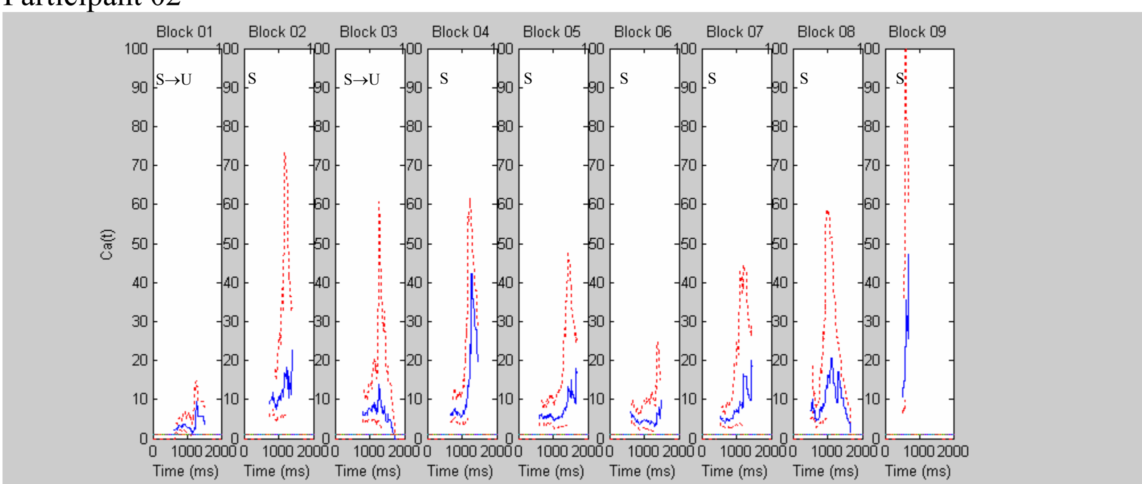


Figure 44: The **Relative learning whole-mixed CCFs**, for the **Sharks members,** across the learning sessions (Block), separately for all participants (blue line). Around each CCF we depict calculated 90% confidence intervals calculated by bootstrapping (red dotted line). We provided the statistical conclusion concerning violations of the C(t)=1 bound, in each small figure. If the CCF was super, unlimited or limited capacity then we used the uppercase letters S, U and L, respectively. An arrow indicates an observed transition between the capacity states.

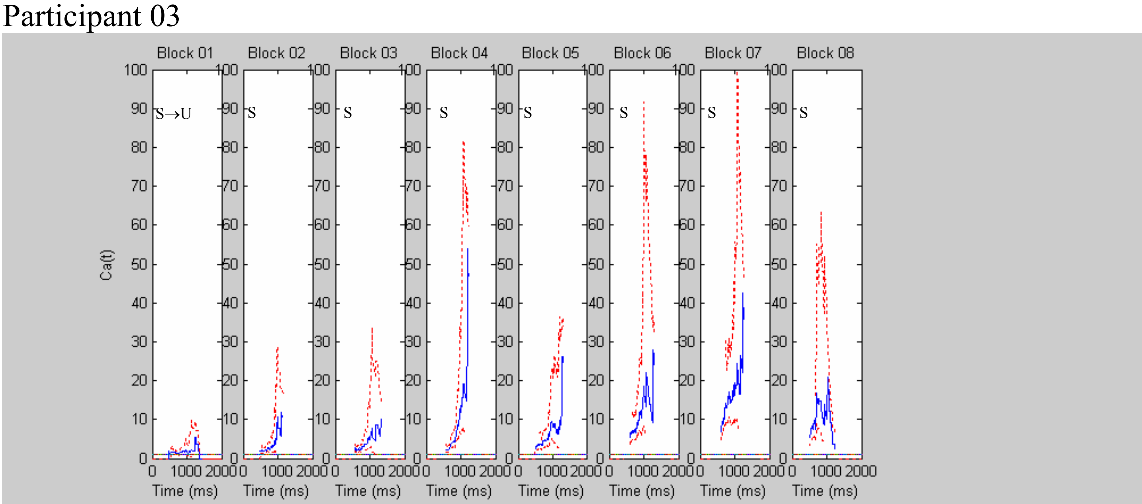
Participant 01



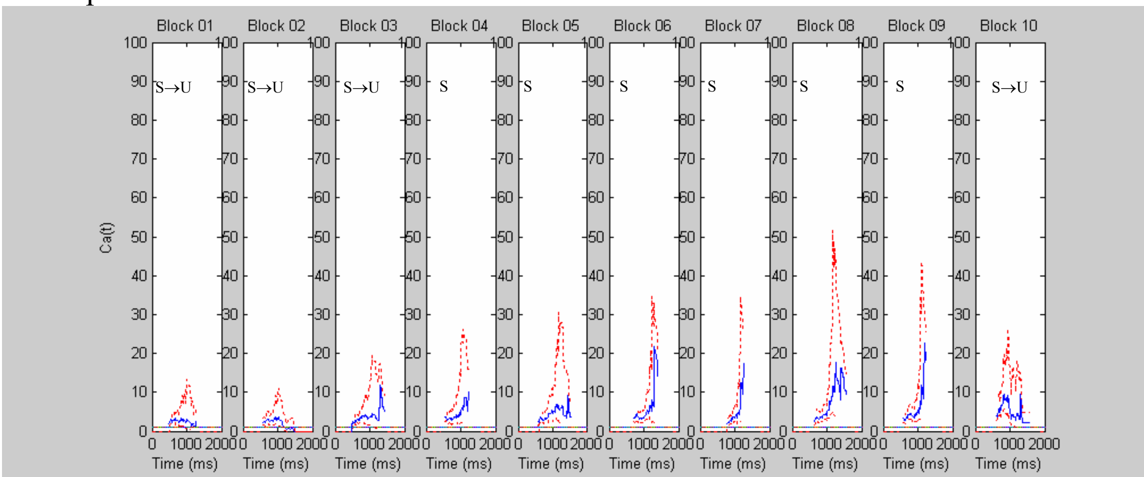
Participant 02



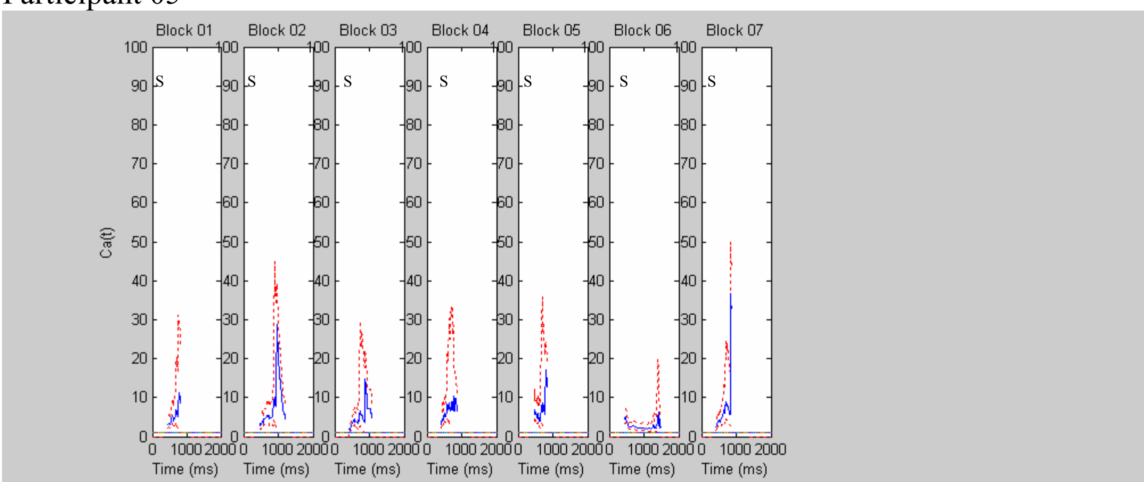
Participant 03



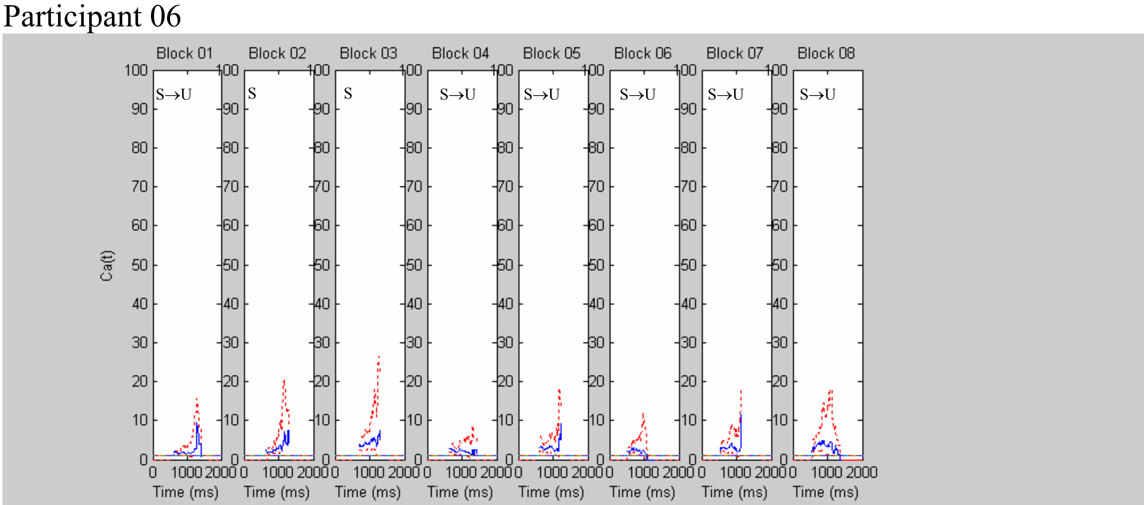
Participant 04



Participant 05



Participant 06



Analyses of Sharks (which allows SFT test)

The overall finding is that most CCFs violate value C(t)=1 toward super capacity. The mixed condition produced lower magnitude CCF values than the blocked conditions. This is consistent with the finding on mean RTs which showed slower processing in the mixed condition. Also, the mixed condition exhibited super capacity CCFs in all cases. Both the absolute and relative learning CCFs clearly violate C(t)=1 toward super capacity in the blocked condition, for almost all participants and learning sessions. Also for some participants, there is a trend of CCF magnitude change.

1. Absolute learning whole-blocked CCF (Figure 41): Several participants exhibited a monotonic increasing magnitude as a function of learning session (Participants 1, 2 and 3 to some extent). Other participants exhibited either a steady stable (Participant 4) or irregular trend (Participants 5 and 6). In all cases, the first block CCF is of smallest magnitude.
2. Absolute learning whole-mixed CCF (Figure 42): Similar to (1), participants exhibited super capacity CCFs with either an increasing trend of CCF magnitude (Participants 1,2, 6) or some non-monotonic trend. The resulting CCFs resembled the ones in (1) with smaller magnitude. Participant 2 exhibited extreme super capacity behavior in the last sessions which did not fit the scaling window; for the sake of scaling issues we leave it blank in the figure and note that its real value is above 100. Careful examination of both confidence intervals for Participant 4 revealed that super capacity was achieved for some short period of time, although in the figure it does not appear so.

3. Relative learning whole-blocked CCF (Figure 43): In this condition, all participants exhibited super capacity for all learning sessions. No monotonic trend of magnitude change as a function of sessions is directly evident, across different participants. We can conclude that for some participants learning occurred at different levels for different sessions. For Participants 3, 4 and 6, it seems that the learning effects were strongest at the beginning and in the middle of the learning sessions, followed by some decrease in learning activity.
4. Relative learning whole-mixed CCF (Figure 44): Super capacity has been demonstrated in all cases, with irregular trends over the learning sessions. As in previous cases, we can conclude that in the mixed condition participants were hampered to benefit from observation of whole-faces only.

Overall, super capacity was observed in all conditions with the largest magnitude in the blocked conditions in both the absolute and relative learning CCFs. It is interesting to observe that participants who exhibited smaller magnitude CCFs in the absolute blocked learning condition showed sizable CCF values in the relative learning blocked condition (Participants 3 and 4). This finding validates the use of different CCF definitions, and will be discussed later. The exhibited super capacity property suggests the possible presence of positive interdependence between face features, which implies the possible presence of gestalt properties during recognition of the Shark faces. Absolute and relative learning CCFs did not show same trends over the learning sessions.

Experiment test phase AND condition

Method

Participants

The same six participants, 4 females and 2 males were paid for their participation.

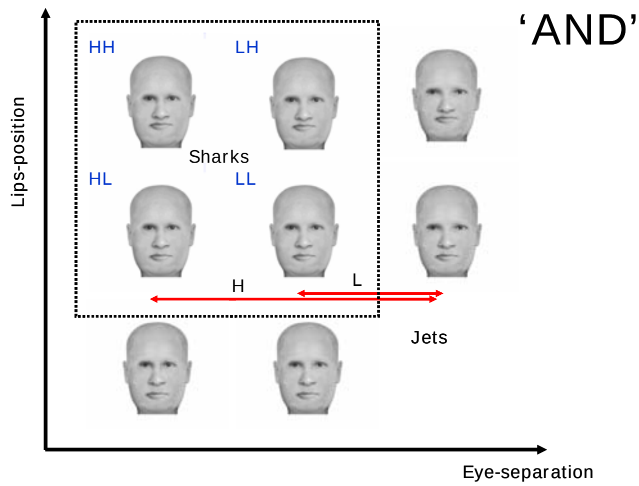
Materials

We used the same materials and design as in the OR test phase experiment. We will briefly remind the reader.

This experiment phase was divided into three different subexperiments: (1) Standard-face test session, identical to the one from the learning phase; (2) Configural-face test and (3) Feature-face test (see Figure 9). Each participant performed each subexperiment 4 times, and order of work for each participant was counterbalanced using the Latin-square design.

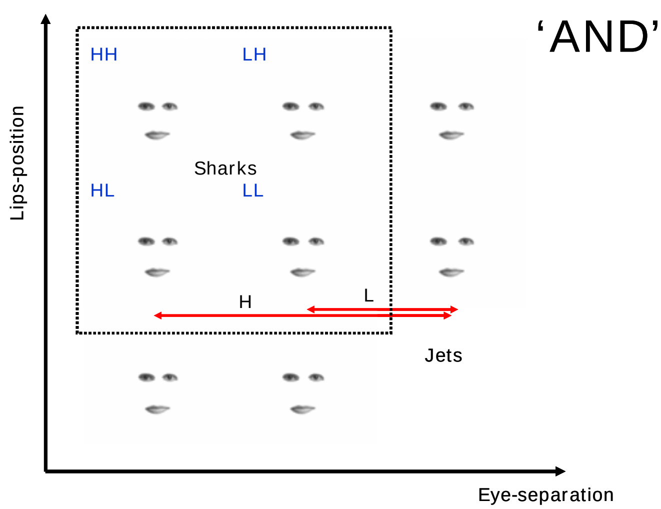
In the configural-test experiment (2) (Figure 45) participants were told following story: “After the incident that happened with the Jets and Sharks, the members from both gangs are hiding from the police. We are informed that they put some disguises on their faces in order not to be recognized. However ALL of them wear the same disguise. The disguise covers everything except the eyes and lips. Have in mind that lips position and eyes separation are the same as before because the disguise does not cover them” (Figure 45).

Figure 45: T**he AND condition, new faces in the configural-test**: two dimensional face-space defined by the eye-separation and lips-position. The saliency of features is defined by the marginal proximity of each face-feature projection onto the axes with respect to the other group members. Boxed faces belong to Sharks, while faces outside the box are Jets. Jets faces are positioned such that they share one face property with Sharks. However in order to recognize a Sharks member, both face features must be recognized. Jets could be recognized on completion of only one feature that is unique for them. In this example Jest posses one feature that is spread-out the most, while all Sharks faces appear more compact. The design is the same as in the standard-test or the learning sessions except that the old faces have their face background completely replaced.



And also, in the featural-test experiment (3) (Figure 46) participants were told following another story: "After the incident they made the members from both gangs are hiding from the police. We are informed that they put some disguises on their faces in order not to be recognized. The disguise covers everything except the eyes and lips. In this session some of the faces presented will have the eyes and lips only! Have in mind that the lips-position and eye-separation are the same as before because the disguise does not cover them". So, in both the featural- and configural-test subexperiments, it was emphasized that the critical configuration that was both necessary and sufficient to generate a correct response did not change.

Figure 46: T**he AND condition, feature-faces in the featural-test**: two dimensional face-space defined by the eye-separation and lips-position. The saliency of features is defined by the marginal proximity of each face-feature projection onto the axes with respect to the other group members. Boxed faces belong to Sharks, while faces outside the box are Jets. Jets faces are positioned such that they share one face property with Sharks. However in order to recognize a Sharks member, both face features must be recognized. Jets could be recognized on completion of only one feature that is unique for them. In this example Jest posses one feature that is spread-out the most, while all Sharks faces appear more compact. The design is the same as in the standard-test or the learning sessions except that the old faces have their face background completely removed.



Design and procedure

Same as in the OR condition.

Results

Basic Mean RT Analyses

First, we compared mean processing time for different subexperiments, the configural, featural and standard tests, for all trials in the experiment and for the Sharks blocked trials. All participants exhibited the following ordering: $RT_{standard}<RT_{feature}<RT_{configural}$ (Table 8, left panel all trials). The main effect of the type of experiment was significant, for all participants, at the level of $p<0.01$, with power=1. So, all participants were fastest in the standard experiment that used the old faces, than in both the new configuration and feature subexperiments.

We also tested the effect of the experiments on processing old faces on the Sharks trials for each participant (Table 8, right panel). Here all participants adhered with the same ordering of means, except Participant 1 who had the following order: $RT_{standard}>RT_{feature}$, ($F(1, 1896)=2.13$, $p<0.05$).

So, we conclude that the standard experiment (3) exhibited on average faster processing than the configural and featural subexperiments where we used configurally altered faces and faces that posses only the features of importance. We can also conclude that new whole faces that are configurally altered mixed together in the same sessions with the old faces affected processing of old whole faces. This is evidenced by comparison of the configural and standard subexperiment (1 and 3). The same holds for the featural manipulation: providing only features in the experiment will alter the processing of old whole faces in the same session when compared to the old whole faces. In both cases, the alteration of faces (featural and configural) produced slower processing of old whole faces. So mixing the configurally altered faces and whole old faces has a detrimental effect on old face encoding, and we suggest that old whole-face processing is directly or indirectly related to the processing of the altered whole-face processing. They either rely on a shared processing mechanism or/and share similar representation space.

Table 8: Mean RTs and standard errors for each subexperimental condition (configural-test, featural-test and standard-test), for individual participants. Table is vertically divided into two parts: all experimental trials were averaged for both Jets and Sharks faces on the left side; and averaged over Sharks faces blocked trials, which are old whole faces from different subexperiments, on the right side.

	All Trials		OLD Shark -members blocked trials	
	Mean	Std. Error	Mean	Std. Error
Participant 01	RT(ms)		RT(ms)	
Configural-test	644.734	2.865	638.818	6.000
Feature-test	652.383	2.863	576.133	5.719
Standard-test	623.265	3.210	591.254	4.698
Participant 02				
Configural-test	711.483	3.908	609.701	6.264
Feature-test	696.500	3.905	576.471	6.124
Standard-test	583.428	4.371	509.194	5.032
Participant 03				
Configural-test	713.643	4.467	681.392	8.942
Feature-test	706.685	4.569	581.906	8.763
Standard-test	633.445	5.103	566.198	7.157
Participant 04				
Configural-test	740.893	4.196	717.490	8.969
Feature-test	747.162	4.266	674.467	8.761
Standard-test	687.851	4.779	641.470	7.169
Participant 05				
Configural-test	566.605	2.457	521.323	4.352
Feature-test	550.662	2.476	486.874	4.310
Standard-test	493.576	2.795	461.245	3.550
Participant 06				
Configural-test	719.726	2.847	686.175	6.258
Feature-test	669.923	2.858	633.130	6.093
Standard-test	660.447	3.216	632.326	5.009

Since we demonstrated that mixing different types of faces can produce changes in the processing of each type (the differences between the standard test and both the configural-test and featural-test), we investigated architectural differences between processing of each type of altered faces (configural and featural) and the processing of old faces in all subexperiments.

Comparison of the processing characteristic differences between old faces and new faces for the standard-test and configural-test (Configural x Standard design)

We ran a GLM analysis, type I model, for separate subexperiments (1 and 3) using the following fixed factors: the eye-separation (high/low), lips-position (high/low), and the experimental group (3 levels: group1=old-configural, group2=new-configural and group3=old-standard). The old-configural conditions are based on trials when old faces from the configural-test were used. The new configural conditions consisted of new faces made from the old faces by configural alteration (changing the face context), also from the configural-test. The old standard faces are taken from the standard-test.

As a covariate, we chose the trial order. Of most interest in the study is the relation between group2 (new-configural faces) and groups 1 and 3 (old-faces). In both groups 1 and 3 we examined processing of whole faces, but the difference is that in group3, whole-faces were not mixed with configurally altered faces. Therefore, we expect that the mean face processing should be fastest in group 3.

Table 9: GLM univariate analysis was conducted on the Sharks faces, for different participants (SUB). The main effects and interaction terms are listed in the second column (*Factor*). Degrees of freedom are in the third column (*df*). The *error* row defines the degrees of freedom for the F-test error term, for that participant. Each F-test value (*F*) has two degrees of freedom: one from its corresponding row, and the second one from the error row. A significance level is presented in the column *Sig.*, and the observed power for that effect is in the last column. Table is vertically divided between two different analyses: a comparison of the processing characteristics between old faces and new faces for the standard- and

configural-tests (Configural x Standard design, on the left side) and a comparison of the processing characteristic differences between old faces and new faces based on removing the face context (Featural x Standard design, on the right side)

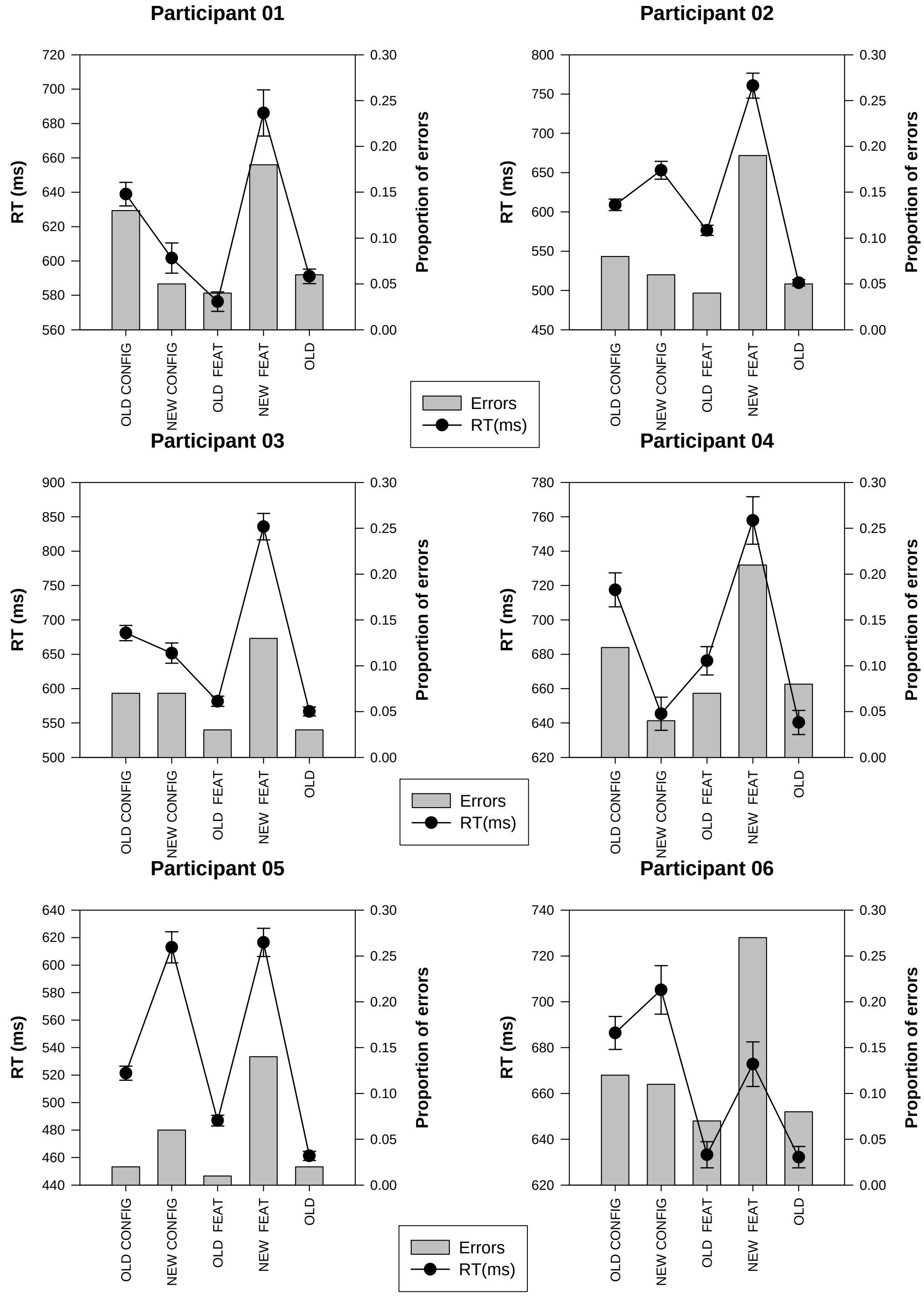
Configural x Standard	df	F	Sig.	Observed Power	Featural x Standard	df	F	Sig.	Observed Power
Participant 01									
Trial order	1	1.125	.289	.185	**Trial order**	1	12.439	.000	.941
Eyes	1	130.218	.000	1.000	**Eyes**	1	83.785	.000	1.000
Lips	1	385.439	.000	1.000	**Lips**	1	436.880	.000	1.000
Exp group	2	31.225	.000	1.000	**Exp group**	2	70.992	.000	1.000
Eyes x Lips	1	5.763	.016	.670	**Eyes x Lips**	1	5.987	.014	.686
Eyes x Exp group	2	9.114	.000	.976	**Eyes x Exp group**	2	1.021	.360	.229
Lips x Exp group	2	9.860	.000	.984	**Lips x Exp group**	2	16.627	.000	1.000
Eyes x Lips x Exp group	2	.429	.651	.120	**Eyes x Lips x Exp group**	2	12.933	.000	.997
Error	2197				**Error**	2213			
Participant 02									
Trial order	1	2.802	.094	.387	**Trial order**	1	.089	.765	.060
Eyes	1	174.895	.000	1.000	**Eyes**	1	163.494	.000	1.000
Lips	1	384.190	.000	1.000	**Lips**	1	224.756	.000	1.000
Exp group	2	161.488	.000	1.000	**Exp group**	2	311.117	.000	1.000
Eyes x Lips	1	13.448	.000	.956	**Eyes x Lips**	1	3.577	.059	.472
Eyes x Exp group	2	3.217	.040	.616	**Eyes x Exp group**	2	3.183	.042	.611
Lips x Exp group	2	10.849	.000	.991	**Lips x Exp group**	2	3.150	.043	.606
Eyes x Lips x Exp group	2	.492	.612	.131	**Eyes x Lips x Exp group**	2	3.438	.032	.646
Error	2238				**Error**	2215			
Participant 03									
Trial order	1	.495	.482	.108	**Trial order**	1	202.882	.000	1.000
Eyes	1	146.282	.000	1.000	**Eyes**	1	139.901	.000	1.000
Lips	1	474.727	.000	1.000	**Lips**	1	490.702	.000	1.000
Exp group	2	66.536	.000	1.000	**Exp group**	2	143.852	.000	1.000
Eyes x Lips	1	.039	.843	.055	**Eyes x Lips**	1	.134	.715	.065
Eyes x Exp group	2	6.044	.002	.885	**Eyes x Exp group**	2	2.338	.097	.475
Lips x Exp group	2	15.348	.000	.999	**Lips x Exp group**	2	25.669	.000	1.000
Eyes x Lips x Exp group	2	3.357	.035	.635	**Eyes x Lips x Exp group**	2	5.467	.004	.850
Error	2276				**Error**	2280			

Participant 04									
Trial order	1	6.535	.011	.724	**Trial order**	1	7.288	.007	.770
Eyes	1	146.021	.000	1.000	**Eyes**	1	147.057	.000	1.000
Lips	1	305.001	.000	1.000	**Lips**	1	197.880	.000	1.000
Exp group	2	35.852	.000	1.000	**Exp group**	2	39.653	.000	1.000
Eyes x Lips	1	1.011	.315	.171	**Eyes x Lips**	1	2.322	.128	.331
Eyes x Exp group	2	1.529	.217	.327	**Eyes x Exp group**	2	.951	.387	.216
Lips x Exp group	2	15.555	.000	.999	**Lips x Exp group**	2	1.437	.238	.309
Eyes x Lips x Exp group	2	.343	.710	.105	**Eyes x Lips x Exp group**	2	2.552	.078	.512
Error	2183				**Error**	2152			
Participant 05									
Trial order	1	.003	.953	.050	**Trial order**	1	.003	.953	.050
Eyes	1	76.870	.000	1.000	**Eyes**	1	81.031	.000	1.000
Lips	1	300.112	.000	1.000	**Lips**	1	298.972	.000	1.000
Exp group	2	195.750	.000	1.000	**Exp group**	2	257.749	.000	1.000
Eyes x Lips	1	1.243	.265	.200	**Eyes x Lips**	1	.074	.786	.058
Eyes x Exp group	2	.235	.790	.087	**Eyes x Exp group**	2	6.305	.002	.899
Lips x Exp group	2	10.660	.000	.990	**Lips x Exp group**	2	9.585	.000	.981
Eyes x Lips x Exp group	2	1.242	.289	.272	**Eyes x Lips x Exp group**	2	.049	.952	.057
Error	2320				**Error**	2301			
Participant 06									
Trial order	1	.337	.561	.089	**Trial order**	1	.535	.464	.113
Eyes	1	179.886	.000	1.000	**Eyes**	1	112.201	.000	1.000
Lips	1	347.972	.000	1.000	**Lips**	1	243.427	.000	1.000
Exp group	2	44.277	.000	1.000	**Exp group**	2	14.990	.000	.999
Eyes x Lips	1	.095	.757	.061	**Eyes x Lips**	1	1.269	.260	.203
Eyes x Exp group	2	9.595	.000	.981	**Eyes x Exp group**	2	.234	.792	.087
Lips x Exp group	2	2.966	.052	.578	**Lips x Exp group**	2	3.821	.022	.696
Eyes x Lips x Exp group	2	.891	.410	.205	**Eyes x Lips x Exp group**	2	.155	.856	.074
Error	2156				**Error**	2127			

Results from the GLM analyses are presented in Table 9 (left) and the mean RTs are presented in Figure 47. We expected to find a significant main effect of the experimental group, as well as three way interaction between the two face features of interest and experimental group (Eyes x Lips x Exp group). The interaction would indicate a possible change the architecture the between different experimental groups observed. Namely, we expect that processing in group 3 (old faces from standard-test) could be based on a different mechanism than processing in the group 2 (configurally altered faces). Given that, on average, we had observed differences between the processing of old whole faces when they were presented alone or combined with configurally altered faces (group 3 and group 1), we could expect changes in the observed face processing architecture. In fact, we expect that by altering the face configuration the system switch from faster system (probably parallel) to less optimal processing (maybe serial) under the constraints of new the face configuration.

From the results presented in Table 9 (left part of the table), we see that all main effects of Lips and Eyes were significant at the $p<0.01$ level and power=1. Also, the main effect of the experimental group was significant for all participants, which indicates that there is significant mean RT processing difference between experimental groups. In Figure 47 we present the mean RTs for the different experimental groups. All participants exhibited the following order of mean RTs: $RT_{old\text{-}faces}<RT_{new\text{-}configuration}$. We conclude that changes in face configuration, which is manipulated by changing the face surround, resulted in overall slower processing of faces. When we apply separate contrast analyses between $RT_{old\text{-}faces}$ and $RT_{new\text{-}configuration}$, all differences reached significance, except for Participants 1 and 4 with $p>.241$ level.

Figure 47: The AND condition: mean RTs from the test phase for Sharks faces. The old and new configural faces are from the configural-test; the old and new featural faces are from the featural-test. The old faces are from the standard-test subexperiment. Error bars around the mean RT indicate standard error.



When we ran separate contrast analyses on mean processing times between $RT_{old\text{-}configuration}$ (old faces in the experimental context with configurally altered faces) and

$RT_{new\text{-}configuration}$ faces from the same sessions, then for three participants (2,5 and 6) we observed the expected significant ordering of the means: $RT_{old\text{-}configuration} < RT_{new\text{-}configuration}$; three participants (1, 3 and 4) exhibited a reverse order: $RT_{old\text{-}configuration} > RT_{new\text{-}configuration}$, while Participant 3 exhibited no significant difference at (p=.102). We previously concluded (Table 3) that old faces were processed more slowly when they are combined with the new-configural faces than when they are presented alone. However, the absence of the expected ordering of old and new faces, depending on experimental context, could be produced by utilization of a special face encoding strategy, and we will return to this issue later when discussing all findings.

At this point, we could suggest that changes in configuration produced significant differences in face encoding. When configurally altered faces are compared to the old faces, for each subexperiment we observed that configurally altered faces were recognized more slowly, for all participants. Also, the MIC tests, revealed by the two-way interaction Lips x Eyes, were significant for Participants 1 and 2. Overall MIC values were positive for Participants 1 and 2, while for others the MICs were very close to additivity (MIC=0). However, further analysis of the MIC values, will be broken down for different experimental conditions, for each participant (see Figures 48, 49 and 50).

Now we will turn our attention to the interaction. The three way interaction Eyes x Lips x Exp group indicates whether there is a possible change of architecture for different experimental groups, between differently altered faces. This interaction between the old faces and new configurations was significant only for Participant 3, ($F(1,1533)=7.345$, $p<0.01$), thereby indicating a possible the change of architecture due to the configural changes in the face.

We can conclude that all participants exhibited mainly an increase in mean RT and only one case of architecture change when the configuration of faces changed. A relative exception to this general finding is Participant 4, who did not exhibit any significant change. Participant 1 did not exhibit an overall slowing down on the mean RT level but he show a change in architecture.

From Figure 47 it can be observed that, across almost all participants, there is a clear trend of error reduction in the old faces condition, but not for the both the old-configural and new-configural faces. The main variable of interest for this study is RT and the participants had the goal of reducing their error levels. The error percentage was not analyzed separately, given that for some cases it was very small with no variance.

Overall, we conclude that all participants processed old faces in the standard comparison test faster than new-configuration faces. This finding is the prime evidence for a configural effect in face perception when two face features are exhaustively processed (AND case). When we change the background of the previously learnt face, then on average we people are slower than when recognizing the original old face. The old faces alone are also processed faster than the old-configuration faces; that is the same old faces are recognized more slowly when they are mixed with new faces, which is, again, support for the thesis that participants are adjusting their global processing strategy (generally speaking) that affects recognition of both types of faces. For several participants old faces in the configural-test are recognized slower than new-configural faces in the same test (Participants 1, 3 and 4). We suggest that these participants favored some aspects of face processing that aided the recognition of isolated features or feature configurations in the new faces, while at the same time, hindered their ability to recognize the old faces.

Figure 48: The SFT tests result for the AND **standard-test** condition, for Sharks faces, for all participants. The results are based on all sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.

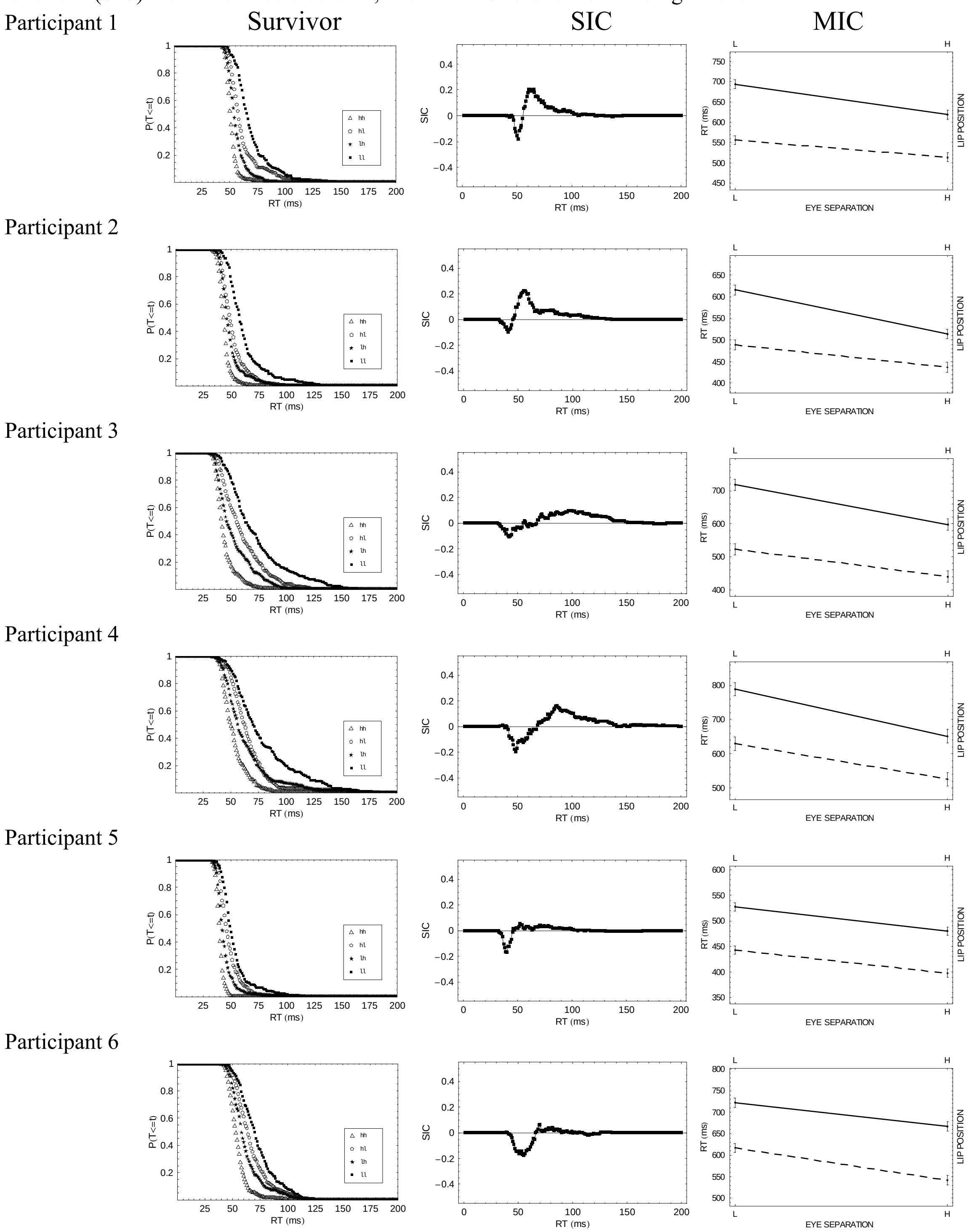


Figure 49: The SFT test results for the AND **configural-test** condition, for Sharks faces, for all participants. The results are based on all sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.

Participant 1 Survivor SIC MIC

Participant 2

Participant 3

Participant 4

Participant 5

Participant 6

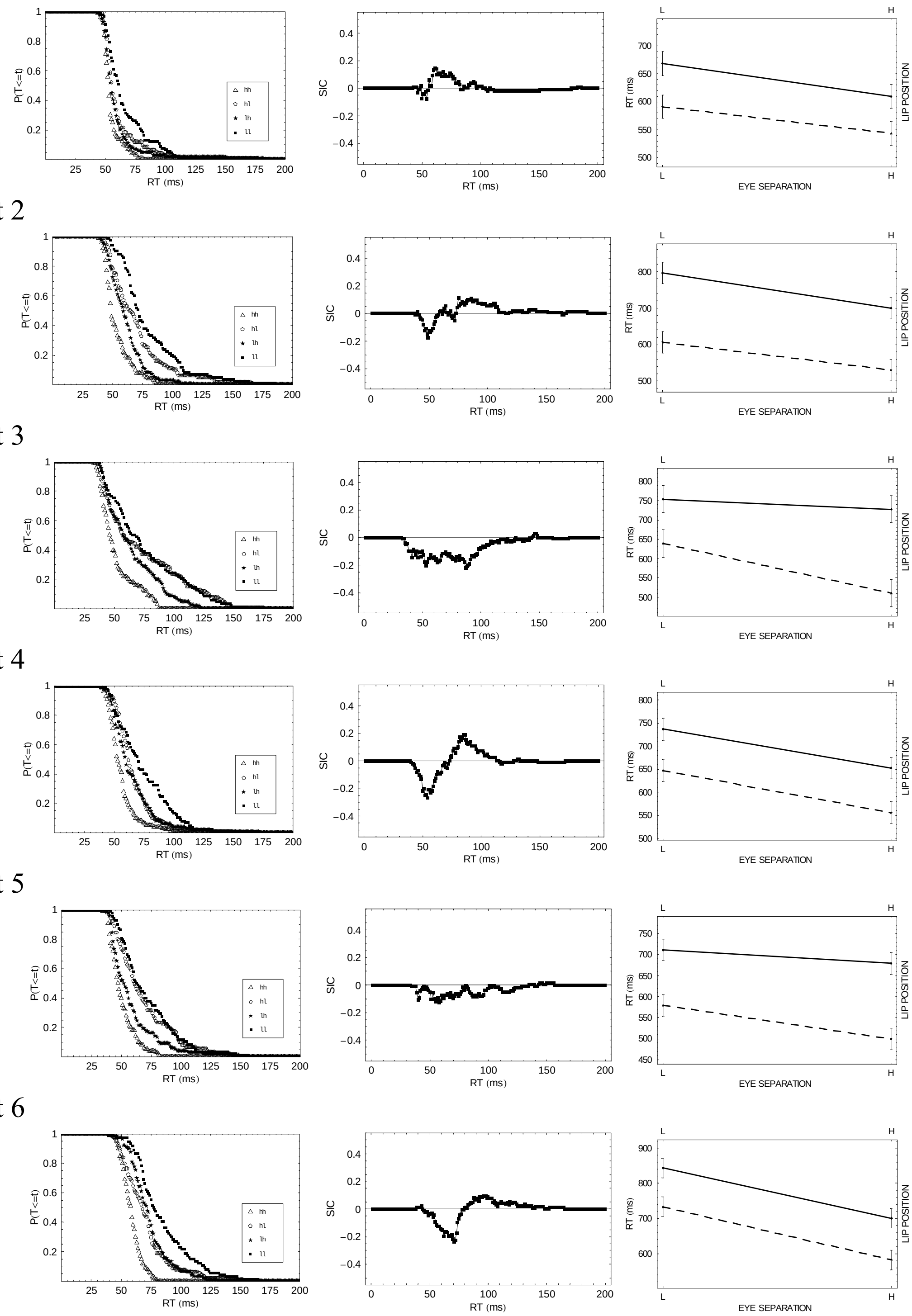
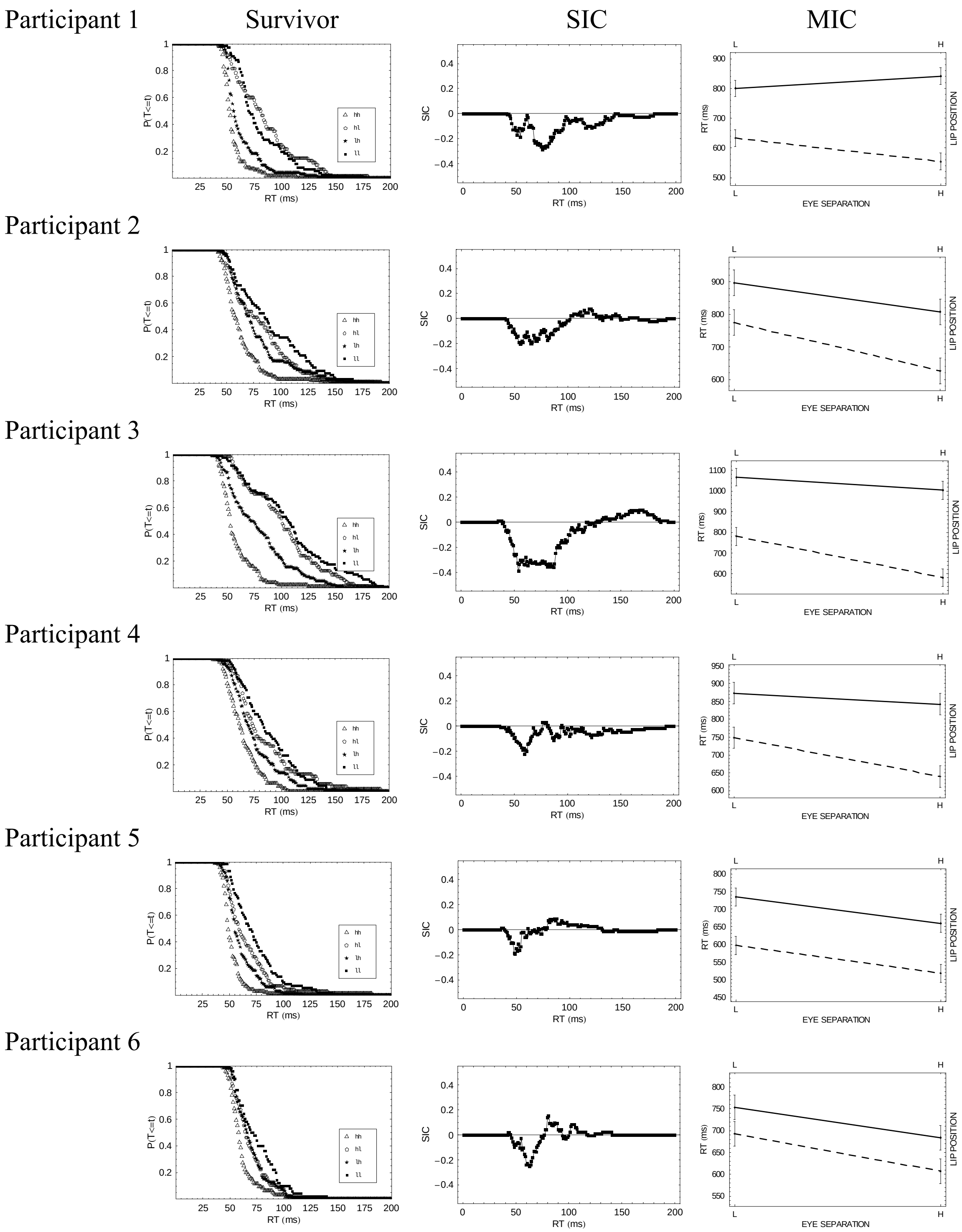


Figure 50: The SFT tests results for the AND **featural-test** condition, for Sharks faces, for all participants. The results are based on all sessions combined. The first column depicts the ordering of the joint survivor functions for the different factorial conditions (HH, HL, LH, LL). The survivor interaction contrast functions (SIC) are in the middle column, and the MIC results are in the right column.



Comparison of the processing characteristic differences between old faces and new faces based on removing the face context (Featural x Standard design)

The same GLM analyses were run on these trials. The only difference was in the data source. The group 1 observations (old configural faces) are collected from the part of the experiment where old whole faces are combined with the feature-only faces (see Figure 9,the 2nd part); The group 2 data stems from the subexperiment but now consists of feature-only faces, and the group 3 data stems from the standard-test subexperiment, with the same whole old faces as in the previous case. The goal of this experiment was to determine the effect of removing the face background, which was not important for the correct decision, on change in mental organization during face processing. Although the face background was not necessary to make a correct decision, if face encoding is based on gestalt/holistic processes, we could expect its removal should affect the configural/holistic strategies, more than the analytic strategies.

The results of the GLM analyses, comparable to the above analyses, are presented in Table 9, right. Also, mean RT for the different experimental groups are presented in Figure 47 (old-featural and new-featural groups) together with the above mean RTs from the configural test. From the results presented in Table 9 (right side of the table) we can see that all main effects of Lips and Eyes were significant at the $p<0.01$ level with power=1. Also, the main effect of experimental group was significant for all participants, which indicates that there is a significant mean RT processing difference between experimental groups. In Figure 47, we present mean RTs for different experimental groups. All participants exhibited a similar pattern of mean RTs: $RT_{old\text{-}faces} \approx RT_{old\text{-}features} < RT_{new\text{-}featural}$, (where old-faces are whole faces from the standard block, old-featural faces are the old faces presented together with the feature-only altered faces, and new-featural are the new feature-only faces (see Figure 47).

Also, the MIC tests results, revealed by the two-way interaction Lips x Eyes, were significant only for Participants 1 and 2. However, further analysis of the MIC values, will be broken down for different experimental conditions, for each participant (see Figures 48, 49 and 50).

The three way interaction, Eyes x Lips x Exp group, indicates whether there is a possibly significant change of architecture for different experimental groups. This interaction was significant for Participants 1, 2, 3 and 4 (for Participants 4 it was marginally significant), and not significant for Participants 5 and 6. A significant three-way interaction indicates a possible change of architecture due to removal of the face background.

From Figure 47 it can be observed that, across all participants, there is a clear trend of error reduction in both the old-faces and old-featural conditions, but not for the new-featural faces. In fact, with the approximate average error level between 15% and 20%, the new-feature condition exhibited the highest error level, compared to all subexperiments, across all participants.

Overall we conclude that there is a significant trend for all participants to process featurally based faces both more slowly and with more errors than the old faces.

Mean and Survivor Interaction Contrast Functions

Both the mean interaction contrast (MIC) and survivor interaction contrast (SIC) test results are presented in Figures 48, 49 and 50, for all three groups of interest:

standard, featural and configural. The results are only for the Shark faces, because the design allowed for only that data to be analyzed with the architecture tests. We present data from each subexperiment.

We focus on the SIC functions presented in Figures 48, 49 and 50. Overall, there are a variety of shapes of SIC functions for different subexperiments and participants. The good thing is that all the SIC shapes are regular and interpretable. In almost all cases the necessary ordering of survivor functions is persevered.

In the standard-test (Figure 48) we can see that Participants 1, 2 and 3 exhibited mostly positive, S-shaped SIC functions. Participant 4 exhibited additive S-shaped SIC function, with a positive area equal to a negative area of the function. Participants 5 and 6 exhibited mainly negative SIC functions. It is important to note that while we would be likely to attribute an S-shaped SIC a serial exhaustive mechanism, given the findings in the learning part, we strongly suggest that this function corresponds to a parallel exhaustive architecture with positively dependent processing units (face features). In the learning phase it was astutely demonstrated that all participants exhibited changes in architecture, while their SIC functions transformed from parallel exhaustive negative SIC functions to S-shaped, mainly positive functions. We concluded that participants in the standard-test condition exhibited similar SIC functions to those exhibited during the learning phase. S-shaped positive functions (with bigger positive part) were taken as a prime evidence of parallel exhaustive processing architecture with dependent processing units. However, there is possibility that some participants switched to serial S-shaped SIC function. Later we will discuss that possibility and show why it is not likely.

The configural-test group revealed similar results at the SIC and MIC levels, with the appropriate ordering of the survivor functions (Figure 49). We can conclude that when faces are configurally altered, that is when their background is removed and replaced with a new background, then the SIC functions remained either S-shaped or negative.

The feature-test SIC analysis revealed uniform findings for all participants: underadditivity, that is negative SIC functions, are evident for most of the participants. This is consistent with parallel exhaustive processing strategy. In contrast to the OR featural-test findings, here all main effects are significant, which suggests that both isolated features produced significant perceptual effects in the AND condition.

We averaged the mean MIC score over different participants and ran repeated measures GLM analysis (Table 10). The main effect between different subexperiments (Table 10) between averaged MICs was significant ($F(2,10)=6.358$, $p<0.05$, power=0.78). The mean MIC value is positive in the standard-test that is based on old whole faces. When the new face background is introduced, then the averaged MIC value is reduced and becomes negative. In the feature-test condition, the averaged MIC value is even more negative that corresponds to mainly negative SIC function. We conclude that the altering of the previously learnt faces produced striking reductions of the averaged MIC values. This finding is uniform over all participants, and we found the same trends when we ran single participant analyses.

Table 10: The mean MIC values and their standard deviations, averaged over all participants, for three subexperiments, on Sharks faces.

	Mean MIC (ms)	Std. Deviation
Standard-test	22.6636	26.20794
Configural-test	-22.1176	46.12938
Feature-test	-69.0931	53.53711

It is interesting to note that the ordering at the mean RT level for almost all participants was $RT_{standard\text{-}test}<RT_{feature\text{-}test}<RT_{configural\text{-}test}$, for the Shark faces (Table 8, right column), except Participant 1 who exhibited $RT_{feature\text{-}test}<RT_{standard\text{-}test}<RT_{configural\text{-}test}$. Similar to the OR condition, it could be expected then that both the SIC and MIC would show some monotonic transition, for example diminishing or increasing magnitude, following the mean RT order. On the SIC level the transitions were not observed between the standard-test and configural-test conditions. But the featural-test condition exhibited change in SIC shape, when compared with both the standard-test and configural-test conditions.

Unlike in the OR condition, the feature-test produced valid MIC and SIC tests, which means that all main effect (lips-position and eye-separation on different levels of saliency) were perceptually distinct and statistically significant. So, when the face background was removed, leaving only two features, processing architecture was clearly parallel exhaustive processing on both the MIC and SIC levels.

The new-face background provided in the configural-test did produced slower recognition of faces on average than in other two conditions. However according to the transition in the MIC and SIC value, it seems that the mechanism achieving this is reducing dependency between processing units.

The second very interesting aspect of the feature test is that, on average, attenuation one face feature (lips) produced faster processing when compared to introducing a new face background (the configural-test). So we can reconstruct the following scenario: removing the face background attenuated one dimension and produced slower processing on average than the standard old faces. Adding a new face background slowed even more, probably because the new face background provided invalid reference cues for spatial detection of features (that could be realized through dependency between face units), but at the same time reengaged on the previously attenuated lips, producing an overall significant effect on them.

If this scenario is true, then the capacity test should reveal smallest CCFs for the featural-test condition when compared with the both the standard-test and configural-test conditions. This was supported by the capacity test results.

<u>The capacity test</u>

We calculated several different capacity coefficient functions for each participant. In Figure 51, we present the calculated capacity coefficient functions, along with bootstrapped 90% percentile confidence intervals, only for Sharks faces (pooled together into one condition). So for each participant, we calculated 4 capacity functions.

We expect that the holistic advantage of processing of whole old faces will be evident as super capacity; that is, the capacity coefficient function will show both confidence interval bounds above the C(t)=1 value. Subsequently, we expect that the CCF for the new faces configural test will be below the C(t)=1 value, for both its confidence interval bounds.

The weaker assumption of a holistic/configural effect does not necessarily predict super capacity effect, but merely a significant difference between CCFs for old faces and new faces. Also given that we found that processing of old faces is sensitive to configural information, we expect that the CCF for old faces will be bigger than the CCF for old faces in the configural test (old faces presented together with the configurally altered faces).

As far as the feature-only condition is concerned, we expect that the feature faces will exhibit CCF confidence intervals that are significantly below the $C(t) \leq 1$ value, revealing unlimited or limited capacity. In the weaker form, the holistic effect assumption predicts that it is not necessary for the $C(t)$ value to be lower then one, because we can expect some configural properties to emerge even from the feature-only presentation, given that some configuration could be inferred from the presentation of the eyes and lips only. So in the weaker form, it is possible that the CCF for feature-only faces exhibits even a moderate upper violation of $C(t)=1$, toward super capacity.

With respect of the amount of learning of the part based information we derived two CCFs. In the first method, the part-based trials of the old faces taken for the calculation stem from the experiment test phase sessions. Implicitly we assume that there is no additional learning conducted on part-based information, starting from the initial learning sessions. In the second method, the part-based trials are from the very first learning session. Note that here we are dealing only with the learning aspect of the part-based information from the old-faces, not the new-faces.

A detailed description of both methods could be taken from the OR condition. The only difference is that in the AND CCF the numerator and denominator reverse their roles.

The capacity test results

We present calculated CCFs for different experimental conditions, for the Sharks faces, in Figure 51 A, for each participant. Both old and new whole face conditions were taken from the blocked sessions, where they were not mixed with part-based information. The first function (top of Figure 51) represents the CCF for old faces from the standard-test condition, where only old faces were presented. In the next figure, the old faces configural test also represents a CCF for old faces, but now taken from the subexperiment when they were presented together with the new faces. The new-faces configural test represents a CCF for new faces from the same subexperiment. The old-faces featural test represents the CCF when old faces are presented together with feature-only faces (see design Figure 9, the 2nd part), and the feature faces from the featural-test is the CCF for feature only faces from the same subexperiment as the previous one.

The right column of Figure 51 B represents the same CCFs calculated by the by the first and second method.

We can observe that for most participants, the old-faces CCF calculated by the second method exhibited the highest magnitude and dominated the other conditions. The exception to this was Participant 4, who exhibited a reverse trend. Also, we can observe that weaker assumption of configural effects is satisfied for almost all cases; that is,

$CCF_{old\text{-}faces} > CCF_{new\text{-}faces}$. All CCFs exhibited super capacity, that includes featural-test faces, significantly violating the C(t)=1 bound.

The configural effect and capacity functions for the Sharks faces

Figure 51 consists of multiple graphs, as described above, for a single participant. Overall, we can observe that all old faces from the standard test (first plot), calculated by both methods exhibited super capacity, for all participants. However, the magnitudes of the CCFs were different for the different cases. In general, the magnitude ordering was established such that $CCF_{old\text{-}faces} > CCF_{new\text{-}faces} > CCF_{feature\text{-}faces}$. All feature faces exhibited very flat CCFs, with the smallest magnitude compared to the other conditions.

In Figure 51 B we report the two graphs of the most interest: the comparison between old faces (calculated by two methods) and new faces. This test is based on comparison of the CCF magnitudes and their corresponding confidence intervals. In order for two CCFs to exhibit a significant difference, both functions must be separated from each other enough that the confidence intervals of the two CCFs do not overlap, at least at some point in time. We also expected the difference to follow a monotonic trend (increasing or decreasing), with no change of sign of that difference observed for different times. That is, we expected that there is no reversal of effects of the expected difference.

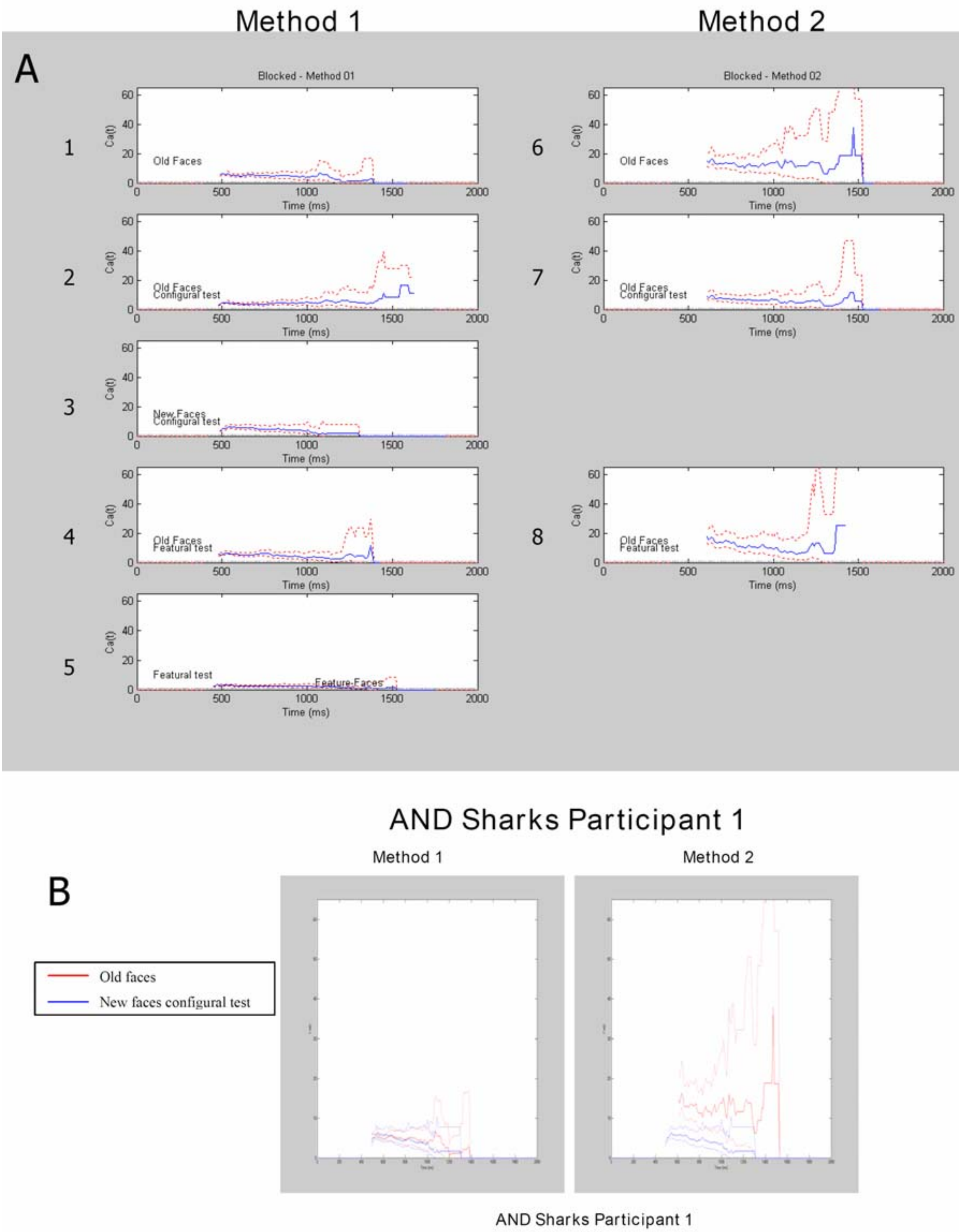
When compared for the configural-test effect, (Figure 51 B) all participants exhibited a significant difference between the old-face and new-face CCFs for method 2. Method 1 of CCF calculation did not show a significant difference between old- and new-faces, for Participants 1, 3, 4 and 6, while it was significant for Participants 2 and 5.

Method 2 established the following order: $CCF_{old\text{-}face} > CCF_{new\text{-}face}$, for all participants, except Participant 4 who exhibited the reverse trend. We will discuss this case later.

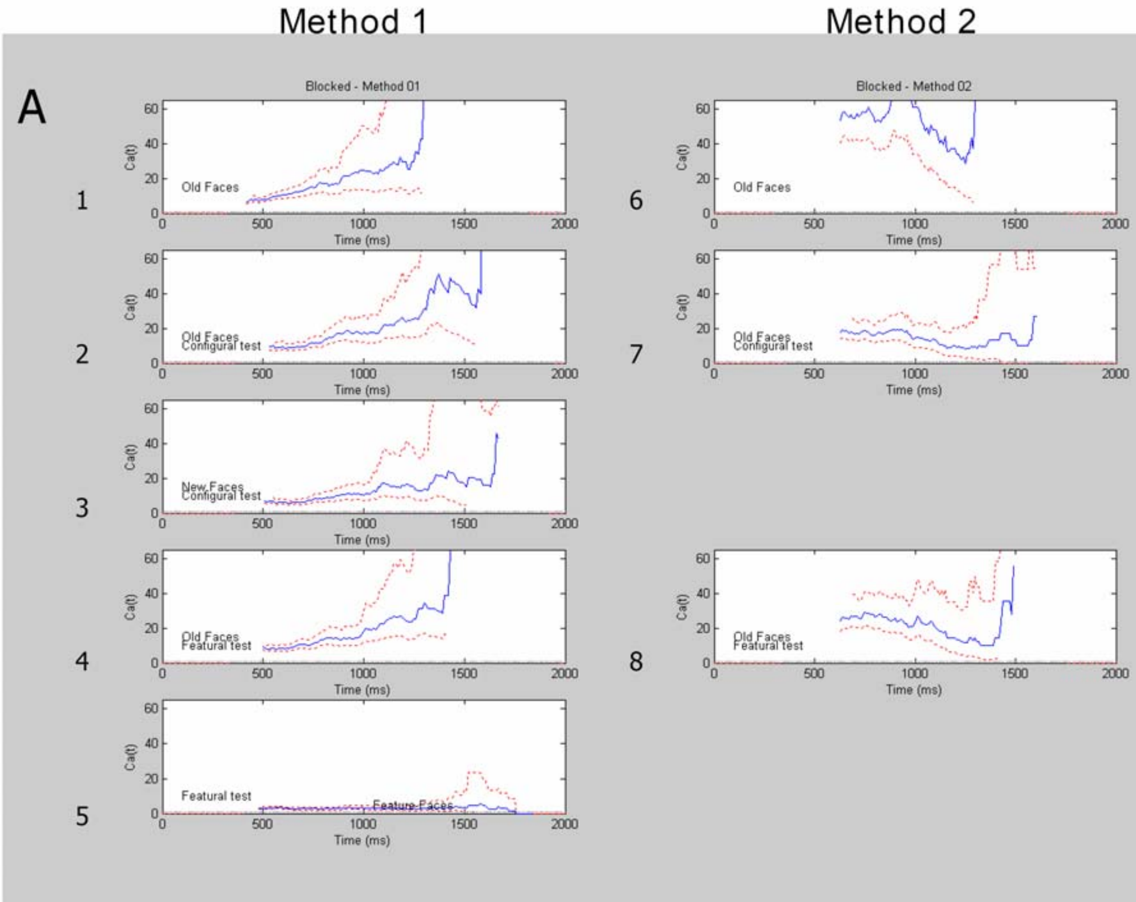
We conclude that configurally altering the old-face into the new-face produced a reduction in CCF magnitude, which could correspond to a reduction in the dependency between face features. All participants exhibited super capacity when processing the new faces. Also, no qualitative change in the architecture was observed between subexperiments. So the introduction of a novel face background was not very detrimental to processing, as we could expect. On the other hand removing the face background (the featural-test) greatly reduced the CCFs magnitude.

The magnitudes of the whole face CCFs were sizable, which could correspond to massive positive interdependency between face features. As in the OR case, we conclude that, in the AND case, face background plays an important role in face recognition detection and aids face detection even if it was not analogous to the background that had been previously learned.

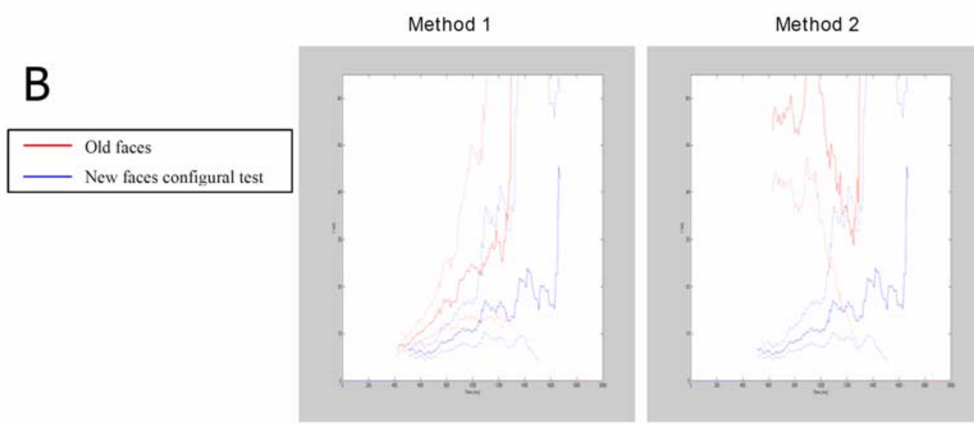
Figure 51: (A) calculated CCFs for different experimental conditions for Shark faces, for each participant, presented on a separate page. Both old and new whole face-conditions were taken from the blocked sessions, where they were not mixed with part-based information. (1) represents the CCF for the old faces from the standard-test condition, where only old faces were presented alone. (2) CCF for the old faces configural test, but taken from the experimental condition when they were presented together with the new configural faces. (3) new faces configural-test represents the CCF for new faces from the same test session as the second CCF. (4) The old faces featural-test represents the CCF when old faces are presented together with feature-only faces in the session. (5) CCF for the feature faces from the featural-test function feature faces only, from that same session as the fourth CCF. The (6), (7) and (8) CCFs are calculated by the second method.
(B) The configural test based on a comparison of bootstrapped confidence intervals, between old faces CCFs (calculated by two methods) and new faces CCFs. Figure on the left is a comparison between (1) and (3) CCFs; and figure on the right is a comparison between (3) and (6) CCFs.



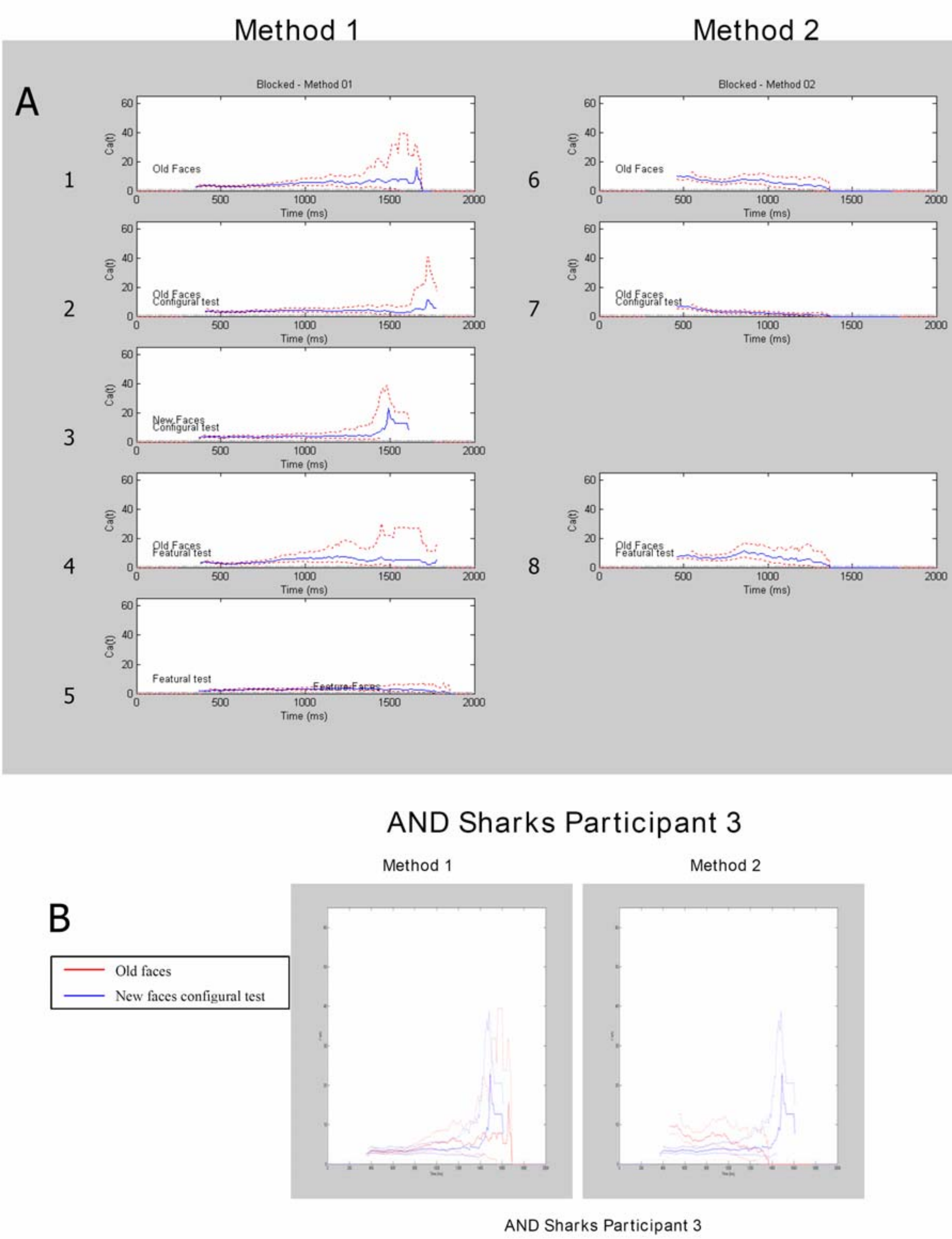
Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Ca(t)
Time (ms)
Old Faces
Old Faces Configural test
New Faces Configural test
Old Faces Featural test
Featural test
Feature-Faces
AND Sharks Participant 1
Method 1
Method 2
B
Old faces
New faces configural test
AND Sharks Participant 1

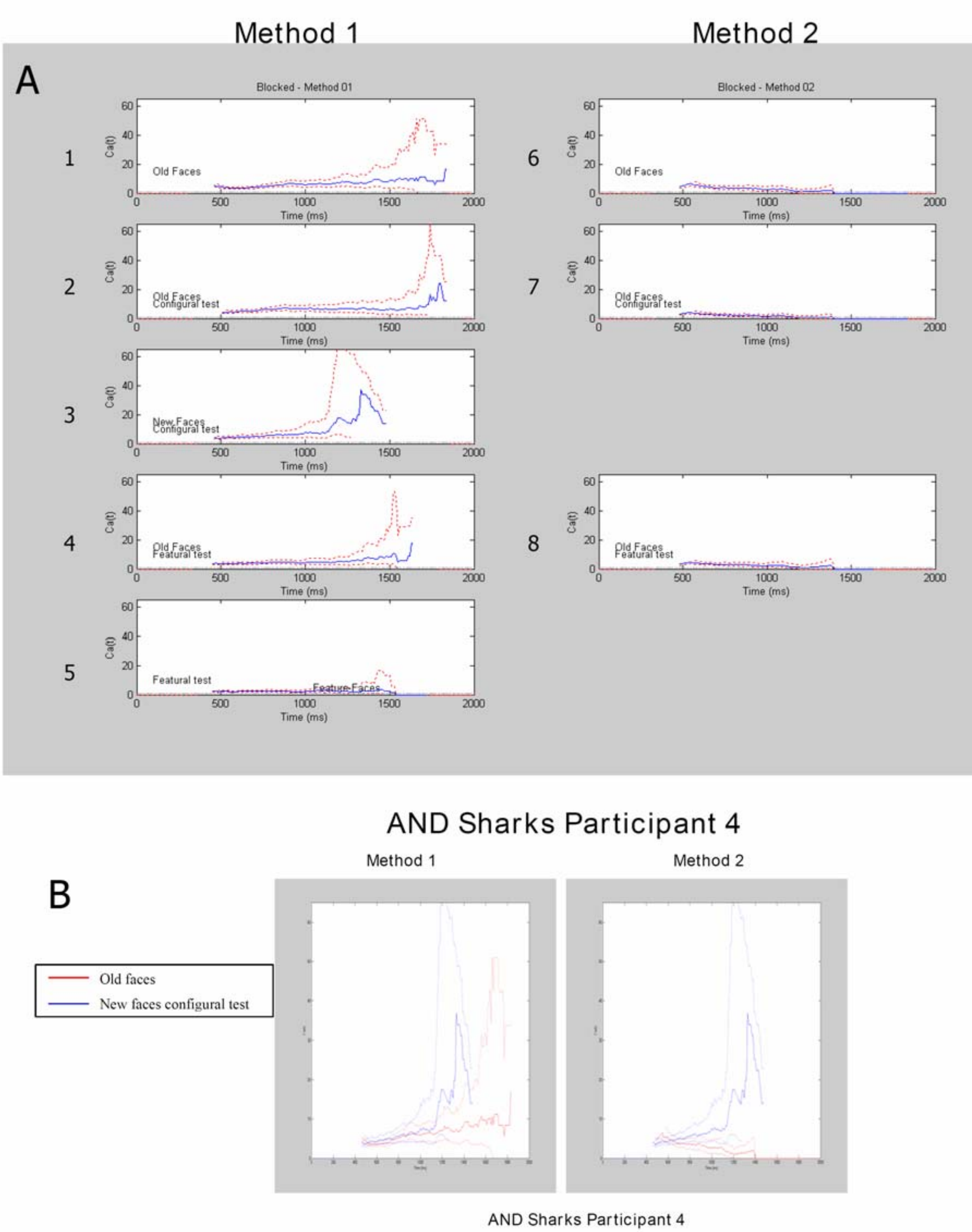


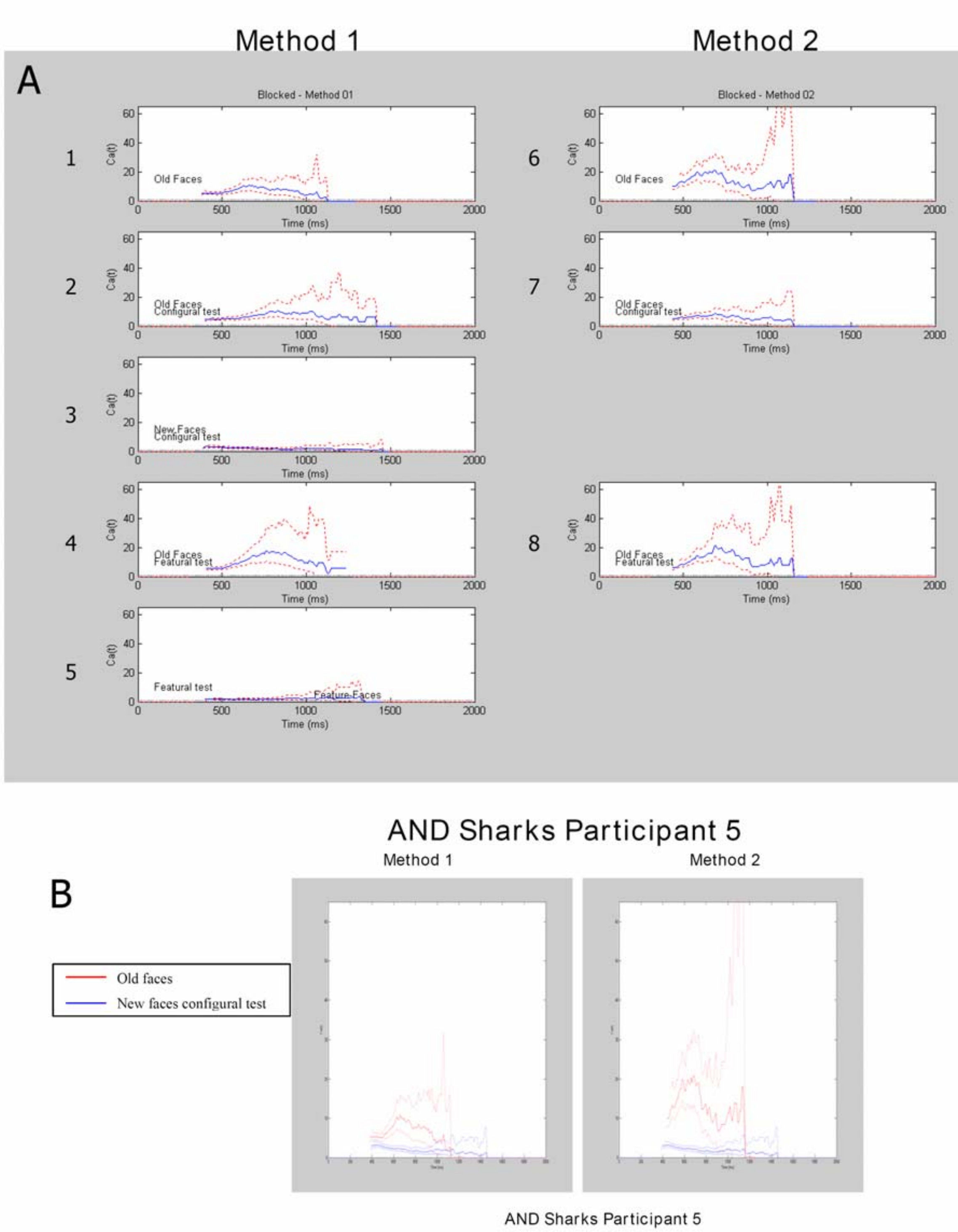
AND Sharks Participant 2

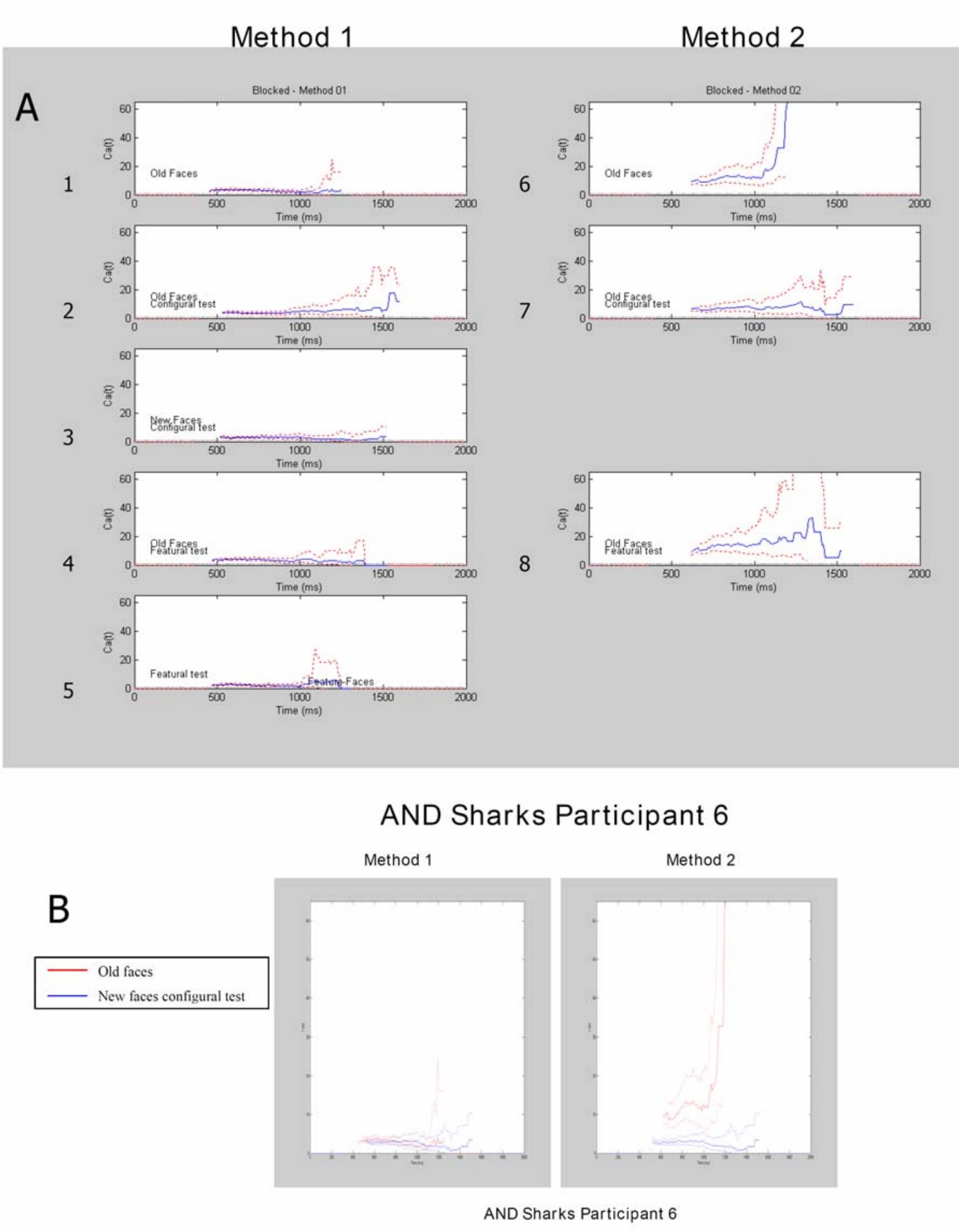


AND Sharks Participant 2


Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Ca(t)
Time (ms)
Old Faces
Old Faces Configural test
New Faces Configural test
Old Faces Featural test
Featural test
Feature Faces
AND Sharks Participant 3
Method 1
Method 2
B
Old faces
New faces configural test
AND Sharks Participant 3


Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Ca(t)
Time (ms)
Old Faces
Old Faces Configural test
New Faces Configural test
Old Faces Featural test
Featural test
Feature Faces
AND Sharks Participant 4
B
Old faces
New faces configural test
AND Sharks Participant 4


Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
1
2
3
4
5
6
7
8
Old Faces
Old Faces
Configural test
New Faces
Configural test
Featural test
Featural test
Feature Faces
Ca(t)
Time (ms)
AND Sharks Participant 5
B
Old faces
New faces configural test
AND Sharks Participant 5



Method 1
Method 2
A
Blocked - Method 01
Blocked - Method 02
Ca(t)
Time (ms)
1
Old Faces
2
Old Faces
Configural test
3
New Faces
Configural test
4
Old Faces
Featural test
5
Featural test
Feature Faces
6
Old Faces
7
Old Faces
Configural test
8
Old Faces
Featural test
AND Sharks Participant 6
B
Method 1
Method 2
Old faces
New faces configural test
AND Sharks Participant 6

General discussion

Holistic face recognition, in broader terms, refers to the ability of human cognitive systems to deal in an integrative manner with separate perceptual units, which we usually call feature representations. An integrated holistic mental representation of a face is not a simple sum of face parts. It possesses unitary properties and corresponds to the whole image/face appearance better than to any individual part of it. Obviously our cognitive system greatly benefits of this holistic quality: when we observe faces in the street we are not cognitively attracted to and aware of their single feature properties. Rather, we see bundles of features as an individual or even a particular identity. It was suggested that the same rules of organization of perception operate in face and non-face object perception (Bruce & Humphreys, 1994; Diamond & Carey, 1986; Tarr, 2003). Other researchers opt for giving special status to the face encoding system (Farah et al., 1995; Farah et al., 2000; Farah et al., 1998; Moscovitch et al., 1997).

In contrast to holistic processing, analytic processing operates at the lower level of perceptual organization. Visual objects are perceived as decomposed parts rather than unitized wholes, and awareness is directed to relatively small details that comprise an object. So, face features, such as nose, eyes, ears, etc., are brought to our perceptual awareness. Analytic recognition is strongly supported by unequivocal findings of the brain's neurophysiological segregation that supports visual scene decomposition (De Valois & De Valois, 1988; Graham, 1989). At the lower levels of the visual pathways such as V1 where columnar organization of neuron cells was discovered, only single features such as differently oriented lines are recognized. Higher neural centers deal with more complex features, usually made of some combination of the simpler features. So with high certainty we state that that a brain decomposes a visual scene and then reconstruct it again. Higher levels of the cortex play a role in what we believe to be holistic recognition (e.g. Grill-Spector et al., 2004; Kanwisher et al., 1997).

Holistic recognition is operationally defined as the advantage for the recognition of a single feature when it is presented in the context of a whole face or whole object (Tanaka & Farah, 1993). The *face superiority* effect has been demonstrated in many studies: face parts are recognized better in their old face context than when they are in a new face context, or when they are presented in isolation (Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999; Leder & Bruce, 2000; Tanaka & Farah, 1993; Tanaka & Sengco, 1997), and in the context of scrambled faces (Homa et al., 1976; Mermelstein, Banks, & Prinzmetal, 1979). This context or part-to-whole superiority effect is not limited to faces but extends to other stimuli such as words, visual objects or geometric (Davidoff & Donnelly, 1990; Enns & Gilani, 1988; Reicher, 1969; Weisstein & Harris, 1974; Wheeler, 1970)

But in realistic face perception, which is based solely on the appearance of whole faces, features are rarely present in isolation. We really do not have opportunity to observe isolated face features. They are always embedded into some face background or face context. On some rare cases, some features could be presented in isolation, such as when doctors wear a face mask. Usually in these cases, our face perception is disrupted to a large extent and other non-face cues are needed to make correct recognition. So the operational definition above is not necessarily at the level at which the strongest holistic properties of face perception are exhibited. Another problem with the part-to-whole paradigm is that it uses face parts mixed with whole faces, and therefore facilitates analytic or part-based information rather than to bolster use of a holistic representation. Presentation of isolated features in the study encourages participants to learn them and to adjust their mental strategies in order to perform optimally. A good question is: do we really change the status of these features presented in isolation? Because single features also possesses good form, it makes sense to assume that the features could be learned as holistic stimuli as well. Then, a possible methodological confound could arise when comparing processing of isolated features and whole faces, especially if they were subjected to different amounts of learning. Some other methods in face studies include only whole faces (e.g. Cabeza & Kato, 2000; Thomas, 2001b; and many proponents of the MDS paradigm). However, using the part-to-whole paradigm could still be beneficial in holistic face-research. But even if we adopt the part-to-whole paradigm, some precautions are necessary. Several theoretical issues were left elusive: how can we describe and characterize both featural and holistic face representations, and how are mental comparisons between the two conducted? For example, let's assume that we formed a holistic representation of John's face, and that his isolated eyes were presented the computer screen. How do we compare that feature with the holistic representation to make a correct decision? What are the processes? How can we compare representations with different formats, at all? These are all problematic issues for the strong holistic hypothesis, which states that faces are unitary representational units.

In the face of findings that support featural component processing, the strong hypothesis of holistic perception is less appealing (as detailed in Farah et al., 1995; Farah et al., 1998; Tanaka & Farah, 1991, 1993; Tanaka & Sengco, 1997). Most of aforementioned questions are left unanswered. However, the proponents of the weaker holistic hypothesis find some shelter. The weaker version assumes that in the holistic representation, not only a configural face representations encoded but also featural representations are encoded. So, featural and configural information are encoded together, and they interact such that when one is changed (manipulated) the other changes as well. Now the weak hypothesis can accommodate variety of part-to-whole findings.

This hypothesis should not be confused with the independent dual-mode hypothesis (Bartlett & Searcy, 1993; Bartlett et al., 2003; Searcy & Bartlett, 1996). In its standard form, the dual-mode hypothesis assumes that both configural/holistic and featural properties are available during face perception, and they operate independently. The difference between weak holistic hypothesis and dual-model is a presence of the interaction between parts and whole face representations. The difference seems subtle especially in the light of the absence of strong mathematical model of face perception. Macho and Ladder (1998) demonstrated how appropriate modeling could be use to better articulate holistic hypotheses.

It is interesting to note that even with the strong holistic hypothesis, Tanaka and Sengco (1997) open the possibility that face parts could be stored separately, given that participants performed above chance during recognition of isolated features. However, that possibility reduces the strong holistic hypothesis to the weak one. This assumption concerning multiple types of face encoding is more general and more flexible. At the same time, it is less likely to be tested and falsified. Particularly if the processes that relate the different representations are not specified, the validity of operational definition of the holistic perception is questionable. It is not surprising, then, that several concerns are raised: Ingvalson & Wenger (2005) argued that none of the studies of holistic face processing involve a logical or formal analysis of the complete set of alternative hypotheses, and use experimental procedure that can allow for strong-inference tests of the hypotheses.

Defining holistic processing in terms of characteristics of processing

Let us consider four general dimensions of real-time information processing: architecture, stopping rule, independence and capacity. By the **architecture** we mean the spatial and temporal arrangement of the psychological processes that are required to perform a given task. Historically, the distinction is between *serial* and *parallel* processing (Atkinson et al., 1969). Serial processing means that all face features are processed in sequential order (for example: eyes first, then lips, then nose, etc.) (Figure 1). Parallel processing means that all face features are processed concurrently – starting at the same time. One special form of parallel processing is coactive processing. Coactivation is defined as a set of parallel processes that converge into a single channel before a logical gate is reached (Figure 2). A logical gate (AND, OR) combines outputs from subsystem operators and defines rules for making a decision.

The second dimension is that of **stopping rule** (Figure 3). A self-terminating rule means that in order to make a decision not all face features are required to be processed. Exhaustive processing means all feature must complete before a decision is made (Colonius & Townsend, 1997; Townsend & Ashby, 1983; Townsend & Colonius, 1997; Van Zandt & Townsend, 1993). For example, a self-terminating search would be to find your friend's face in a crowd only by examining eyes. If you are familiar with her, face then probably only the eyes (for example) could help you successfully search. However, it is possible that such perception is not possible and you might need to process all face features. This would be a case of exhaustive processing.

The third dimension of processing is **independence** (Figure 4). By this we mean the degree to which the processing rate of any one feature or element affects the rate of processing of any or all of the other elements. If there are positive dependencies between perceived features, then recognition of the first will aid recognition of the second one, and vice-versa. In our example, it might be possible that when the eyes are recognized, it speeds up recognition of the nose.

The final dimension is that of process **capacity** (Figure 5). The question of interest is the manner in which a system responds to manipulations of workload. If performance is unaffected with the increase of workload, then the system is believed to have unlimited capacity. Consequently if performance declines, the system is of limited capacity. And if performance improves with an increased load, then the system possesses super capacity. An example of limited capacity would be a case when increasing the number of features in a face yield slower processing. Our cognitive system might take more time because with each added feature there is more work to be done.

Conclusions

In our experiments we manipulated two face features: the eye-separation and lips-position. Both could be categorized as the second-order relational features (Diamond & Carey, 1986). Many studies have demonstrated the importance of spatial relations between face features, which could be stored or encoded in a face representation (Diamond & Carey, 1986; Haig, 1984; Hosie et al., 1988; Rhodes, 1988). At least two features are needed to test the holistic effect. Configural or spatial changes on faces produced a set of four faces (the gang or Sharks depending on the experiment) on which we applied the SFT technology by calculating the MIC, SIC and capacity functions. Participants went through a series of learning sessions in order to establish and to emphasize holistic encoding of these faces.

During the learning phase, according to introspective reports, participants started with some form of an analytic strategy. But after several sessions, faces started to appear as whole individuals, sometimes described by their perceived emotion, social identity or even just as someone's familiar face. Both the OR and AND conditions revealed exactly what was observed at the introspective level: the architecture of the face processing was, in both conditions, parallel exhibiting exactly predictable shapes for each specific termination rule. In the OR condition, parallel self-terminating SIC functions dominated; in the AND condition, initial learning sessions revealed parallel exhaustive search. The strong validity of these findings relies on the strong validity of the SFT technology.

In addition to the SIC tests, the capacity coefficient functions revealed lower magnitudes for initial sessions, though still of super capacity in both the OR and AND conditions. This occurred in the blocked condition for both the absolute and relative learning CCFs. Absolute learning exhibited a trend of increasing magnitude across learning sessions. Some participants did not exhibited an increasing trend (or it was very weak), but for some of them, this was affected by the presence of a speed-accuracy trade-off that dominated during the first learning sessions. The notion of capacity is closely related to the hypothesis of independence between processing of face features; so it can be concluded that face features exhibited some weak dependency at the beginning, yielding super capacity CCFs. Additionally, they exhibited a trend of increasing dependency during the learning phase.

During later learning we observed changes in the architecture. In the AND condition, architecture exhibited a marvelous transition from a negative SIC function to an S-shaped SIC, in many cases with larger positive area than negative. All participants exhibited this trend, some of them to a larger extent than others. Although the S-shaped function with equal areas is typically considered to serial exhaustive processing, it could be implied that the architecture switched from parallel to serial during the learning. However the architecture switch could be rejected based on two criteria: (a) theoretically it is not acceptable that processes that become faster exhibit the change from parallel exhaustive to serial exhaustive processing in conjunction with an increase in the capacity index, and (b) the simplest model of the observed transition is one that supports a change in dependency between processing units, rather than a change in architecture. However, the observed changes in both architecture and capacity can be plausibly explained by a parallel system that develops positive dependency over the course of learning, followed by an increase in the single rate parameters for both channels (features). This explains the data very well in a simple and elegant fashion. Moreover, it was earlier demonstrated by

Townsend and colleagues that some real time system could exhibit the observed behavior with a set of relatively simple assumptions concerning the organization of processing systems (Townsend & Nozawa, 1995; Townsend & Wenger, 2004).

The OR condition did not show such obvious transitions on the SIC level as a function of learning. For most of the participants, data did show increasing trends of absolute and/or relative learning CCFs, except for two participants who exhibited a speed-accuracy trade-off. Overall, we concluded that super capacity increased as a function of learning. That strongly implies an increasing dependency between face processing units over the course of learning.

In fact, both analytic calculations and simulations explain why it is not possible to observe a strong change of SIC as a function of increasing dependency between units in the minimum time parallel processing system. The shape of the SIC function does not change much in its appearance for different levels of positive dependency between processing units. The only indicator that processing has reached the limits of extreme super capacity is a small negative "blip" value of the predominantly positive SIC function. That model is defied as coactive processing system (Figure 2 B). We observed the appearance of several negative blips in the data, but these were mainly associated with middle learning sessions. Regardless of the sizable CCF, we were not able to observe lawful changes in the appearance of the negative blip as a function of the learning, and can not conclude that the parallel self-terminating model switching in a coactive structure by the end of the learning phase. However, we can conclude that the parallel minimum time architecture was preserved and that interdependency increased to a large extent toward the later training sessions.

In order to illustrate these conclusions, we provide a demonstration of the parallel model for both the AND and OR cases as suggested above, and show how through lawful manipulation of the rate parameters of a single face feature and the rates that describe the changes due to learning, we can mimic the observed behavior of the data. Note that the model presented here is a stochastic model, and it is a replica of the dynamic system suggested in the Townsend & Wenger parallel processing system (Townsend & Wenger, 2004). For the sake of simplicity, but without any loss in generalization, the dependency between units has been realized as a violation of selective influence. So we assumed that an increase in each channel rate is accompanied by some increase in the other channel rate, assuming a positive, direct dependency (Townsend, 1984; Townsend & Thomas, 1994).

Modeling

Most of the participants exhibited a uniform pattern of changes in both the SIC and capacity tests over the course of learning. The SIC function showed lawful changes in the shape indicating specific transitions of architecture, especially for the AND condition, while the absolute learning CCFs exhibited increasing trends for most of the participants. To our knowledge, the simplest model that can encompass these findings is a parallel architecture that develops positive dependence between processing units over the course of learning.

The goal of this section is to provide modeling support for the assumption of an emerging property of face holism that was evidenced by our participants during learning. We assume that participants started with an independent parallel model, for both the AND and OR cases. Over the course of learning, the cognitive system develops facilitatory cross-channel structure between face-feature detectors. The larger the facilitation is, the more faces appear holistic.

We will briefly discuss one possible form of interdependence between processing units based on violations of direct selective influence. Selective influence is defined as the 1:1 correspondence between experimental manipulations and internal subsystems. A violation of selective influence means that manipulation of a single experimental factor influences both (or several) subsystems of processing. Mutual facilitation between processing units can occur through the failure of selective influence in a direct way. In our case that could be exemplified as follows: If selective influence holds then, for example, the experimental manipulation on the eye-separation influences only the eye-separation detector, not the lips-position detector. However, when selective influence fails then each experimental manipulation could affect both face-feature detectors, thus producing the dependency effect. This violation of selective influence is defined as a direct dependency (Townsend, 1984).

Selective influence can also bi violated indirectly. For example, if during learning phase the face detectors for eye-separation and lips-position exchange accumulated information concerning the positive recognition of each feature, then processing of the other channel or detector could be speed up. If the system is more certain that the eye-separation belongs to John (for example), then the lips-position detector can benefit, and speed up its processing. This is an example of positive indirect dependence between units of processing (e.g. Townsend, 1984). It is indirect because that the rate of the eye-separation accumulator can affect the rate of accumulation of the lips-position detector, indirectly via the accumulation time of the random variable from the eye-separation detector.

In our simulations, these two forms of interdependency provided similar predictions regarding the shape of the SIC function in both the AND and OR cases. For the sake of simplicity, in our simulation here, we will use the direct non-selective influence dependency system. We will also constrain our approach to the use of an exponential density function as a statistical descriptor of completion time for each feature detector (channel). But, it will not affect the generalizibilty of the demonstration given that we will provide only a qualitative account of our findings.

We will use parallel self-terminating and exhaustive stochastic models to describe the variability of recognition time as a function of learning session. We assume that the completion rates of each of the two face-feature detectors are defined by an exponential density function (for simplicity). We provide models for both the OR and AND experiments.

The general mathematical form of this model is very similar to the dynamic model between two parallel channels (Townsend & Wenger, 2004) where real-time dynamic system is used, based on simulation. In this study, we will derive the analytic stochastic forms of that dynamic model, for both the OR and AND conditions. Our goal is to demonstrate that the simplest possible model can qualitatively account for the exhibited data. We try to account on the changes for the SIC functions, mean RTs and CCFs as a function of learning session.

OR case, parallel self-terminating model

The joint density function of the parallel self-terminating (minimum time) independent model, on two processing units can be written as (Townsend & Ashby, 1983):

$$f(t;x_{eyes},x_{lips}) = f_{eyes}(t;x_{eyes})\cdot S_{lips}(t;x_{lips}) + S_{eyes}(t;x_{lips})\cdot f_{lips}(t;x_{lips})$$

A marginal density function is denoted with *f(t,x)*, and a marginal survivor function is *S(t,x)*, for all *t={0, ∞}*; rate parameter of information accumulation is denoted by *x*. Thus, x_{lips}, x_{lips} are the distribution rate parameters that define the speed of accumulation for each face feature (eyes and lips). The two processing face units refer to eye-separation and lips-position. The corresponding joint survivor function is:

$$S(t';x_{eyes},x_{lips}) = \int_{t=t'}^{\infty} f_{eyes}(t;x_{eyes})\cdot S_{lips}(t;x_{lips}) + S_{eyes}(t;x_{eyes})\cdot f_{lips}(t;x_{lips})\ dt$$

We assume the dependent processing parallel minimum-time architecture with direct non-selective influence. The joint density function and the survivor function are defined as:

$$f(t;x_{eyes},x_{lips},modif_1,modif_2) = f_{eyes}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_1})\cdot S_{lips}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_2}) + S_{eyes}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_1})\cdot f_{lips}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_2})$$

$$S(t';x_{eyes},x_{lips},modif_1,modif_2) = \int_{t=t'}^{\infty} f_{eyes}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_1})\cdot S_{lips}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_2}) + S_{eyes}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_1})\cdot f_{lips}(t;\frac{x_{eyes}\cdot x_{lips}}{modif_2})\ dt$$

Note that the direct non-selective influence is realized through the multiplication of parameters x_{lips}, x_{lips} between processing units. Thus both parameters appear in each density function. The modifiers of rate parameters are denoted as $modif_1, modif_2$ and are defined in a similar way as the cross-talk parameters in the dynamical systems model of Townsend and Wenger (2004).

The role of modifier parameters is to affect the values of each unit rate parameter x_{lips}, x_{lips}, and therefore affect the speed of processing of corresponding face unit. The dependent parallel minimum-time model, from equation above, could be reduced to independent if we set the modifiers such that $modif_1 = x_{lips}$ and $modif_2 = x_{eyes}$ and replace them in above equation. The magnitudes of $\frac{1}{modif_1}, \frac{1}{modif_2}$ defines the level of mutual interaction between processing units. If the value of that ratio is getting larger then the strength of positive dependency is larger as well.

We will also assume the following constraints:

$$x_{lips}, x_{lips} = \{high, low\}$$
$$1 \leq x_{lips}, x_{lips} \leq 0$$
$$modif_1, modif_2 \leq 0$$
$$1 \leq high, low \leq 0$$
$$low \leq high$$

<u>SIC test implementation</u>

We will provide an example of how both the SIC function and mean RT change as a function of learning (sessions). Recall that the SIC function is defined as the double difference between the 4 joint survivor functions for the double factorial conditions, where each subscript letter denotes either the high or low salience level of manipulation of a single processing unit.

$$SIC(t) = S_{ll}(t) - S_{hl}(t) - S_{lh}(t) + S_{hh}(t)$$

We assume that the rate parameters x_{lips}, x_{lips} are constant over learning, and that the modifiers will *monotonically* change their values over the course of learning. We will also assume that the first session exhibits no cross-talk between channels, which is equal to the parallel independent self-terminating system. Table 11 below lists the values of parameters used in the calculation, where each row corresponds to one simulated learning sessions (7 sessions).

The values of the parameters $modif_1, modif_2$ are further constrained such that they are not free to vary over the learning sessions. We assume that their values will change as a function of learning according to the following equation:

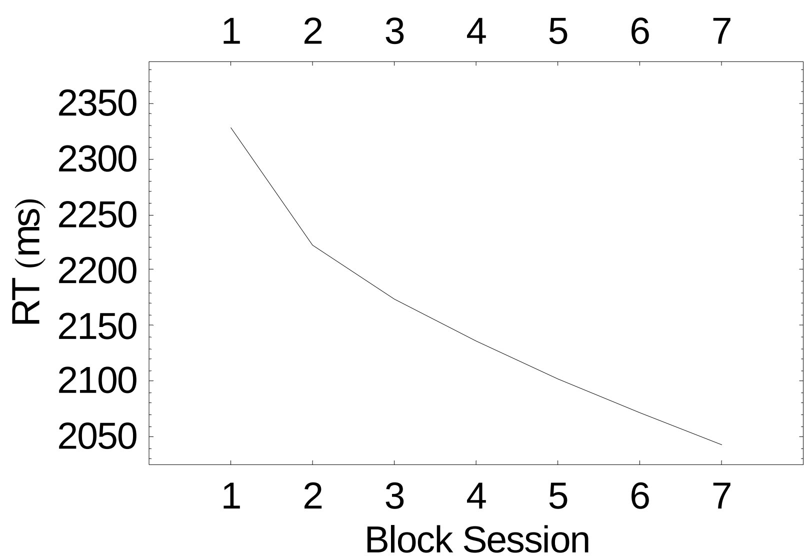
$$\text{modif1}=0.0013 - 0.000191118 \cdot \sqrt{\text{Session}}$$
$$\text{modif2}=0.0005 - 0.0000024711 \cdot \sqrt{\text{Session}}$$
where Session={0,1,2,3,4,5,6}

The value of *Sessions* was zero for the first learning session in order to achieve that the modifiers have the same value as the corresponding rate parameters for each processing unit.

Table 11: Values of the parameters used in simulations for both the OR and AND experiments.

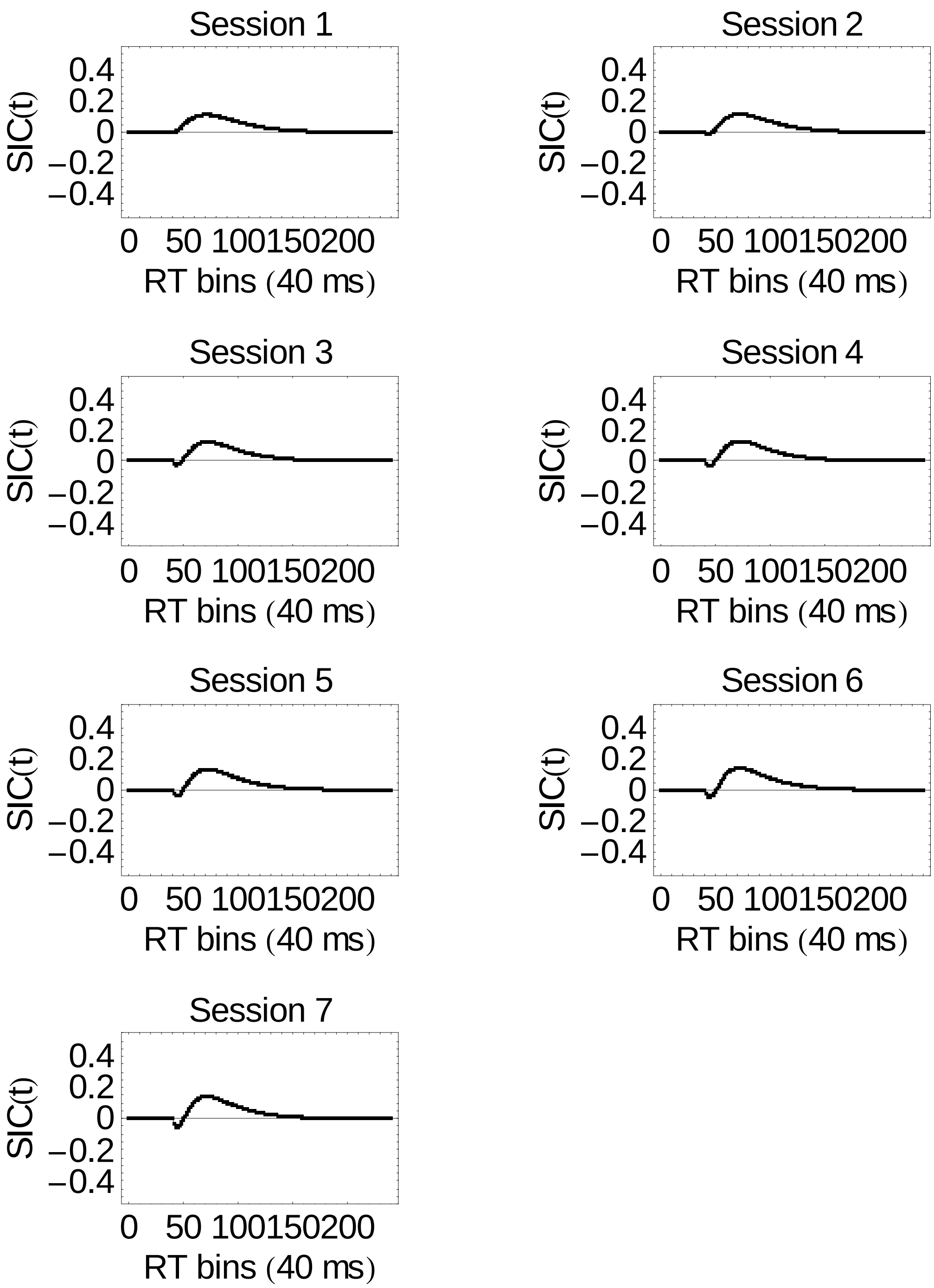
	high	low	$\frac{1}{\text{modif1}}$	$\frac{1}{\text{modif2}}$
Session 1	0.013	0.005	76.9231	200.
Session 2	0.013	0.005	90.1809	210.398
Session 3	0.013	0.005	97.114	215.029
Session 4	0.013	0.005	103.202	218.723
Session 5	0.013	0.005	108.961	221.937
Session 6	0.013	0.005	114.594	224.848
Session 7	0.013	0.005	120.213	227.547`

Figure 52: Calculated mean RT over learning sessions. The trend of RT reduction qualitatively corresponds to the observed data on each group of faces in the OR experiments.



In Figure 53, we present the SIC function over different learning sessions. As we can see from Figure 53 the shape is primarily overadditive during different stages of learning. It is interesting to observe that the appearance of the negative blip, a small negative deviation from the overall positive SIC function, was exhibited for fast reaction times. The magnitude of the negative blip tends to increase as a function of the learning session.

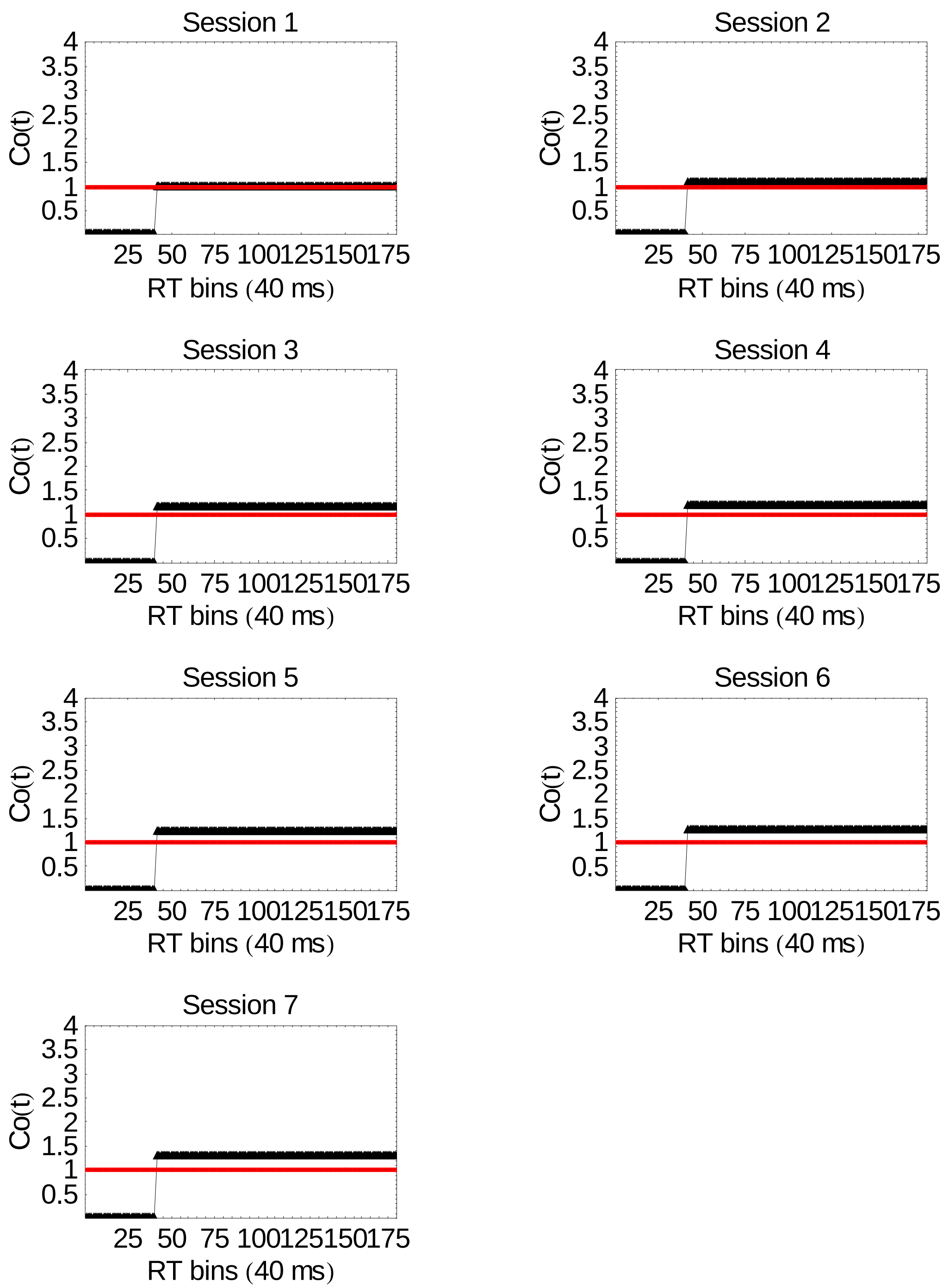
Figure 53: Simulated SIC functions for parallel positively dependent minimum-time (OR) processing model, across learning sessions.



Absolute learning CCFs

In Figure 54, we present the absolute learning CCFs over different learning sessions. Recall that in the absolute learning CCF, the integrated hazard ratio of the processing of parts, from the denominator, has the same value over learning session. In order to characterize the processing of face parts (in the denominator) we arbitrarily set their respective values to $\{x_{lips}, x_{lips}\}/2$, for each face feature. As we can see from the figure 54 the magnitude of the CCF is around one in the first session, and tends to increase in value with the increase in positive facilitation between units. This simulation provides, qualitatively an account of our experimental findings. Note that the flat shape of the CCF is a simplification artifact due to utilization of the exponential density function, and should not be considered important.

Figure 54: Simulated the absolute capacity coefficient functions (CCFs) for parallel positively dependent minimum-time (OR) processing model, across learning sessions.



Discussion

We conclude that mere manipulation of the intensity of the cross-talk structure, realized through the modifying parameters, provided an appealing approach to explaining data. All aspects of analysis mean RT, overall shapes of the SIC functions and CCFs showed very similar trends in simulation as functions of learning. For example, observe the results of Participants 3, 5 or 6, in the OR condition. However, the simulation showed an evidence of the coactivation in processing: a size of negative blip (small deviation from mostly positive SIC functions) tends to increase with increasing positive dependency between face feature detectors. For most participants the real experimental data did not exhibit a monotonically increasing small negative blip deviation. On the other hand, the results of Participants 5 or 6 did show an increasing small negative blip, but which rarely appears in the later sessions, although they exhibited monotonically increasing the absolute CCFs. A possible explanation is that we have not included noise in the model above, so it is possible that the noise can mask the negativity in some cases. However we also leave open the possibility that some other factors might influence their disappearance in the real data, that can include sampling variability.

AND case, parallel exhaustive model

The joint density function of the parallel self-terminating (minimum time) independent model, on two processing units could be written as (Townsend & Ashby, 1983):

$$f(t_{eyes}, t_{lips}; x_{eyes}, x_{lips}) = f_{eyes}(t_{eyes}; x_{eyes}) \cdot S_{lips}(t_{eyes}; x_{lips}) \cdot f_{lips}(t_{lips}; x_{lips}) + f_{lips}(t_{lips}; x_{lips}) S_{eyes}(t_{eyes}; x_{lips}) \cdot f_{lips}(t_{lips}; x_{lips}) \cdot$$

A marginal density function is denoted with *f(t,x)*, and a marginal survivor function is *S(t,x)*, for all *t={0, ∞}*; rate parameter of information accumulation is denoted by *x*. Thus, x_{lips}, x_{lips} are the distribution rate parameters that define the speed of accumulation for each face feature (eyes and lips).

Next, we assume the above architecture dependent processing through direct non-selective influence. The joint density function and survivor functions are presented as:

$$f(t_{eyes}, t_{lips}; x_{eyes}, x_{lips}) = f_{eyes}(t_{eyes}; \frac{x_{eyes} \cdot x_{lips}}{modif_1}) \cdot S_{lips}(t_{eyes}; \frac{x_{eyes} \cdot x_{lips}}{modif_2}) \cdot f_{lips}(t_{lips}; \frac{x_{eyes} \cdot x_{lips}}{modif_2}) + \\ f_{lips}(t_{lips}; \frac{x_{eyes} \cdot x_{lips}}{modif_2}) S_{eyes}(t_{eyes}; \frac{x_{eyes} \cdot x_{lips}}{modif_1}) \cdot f_{eyes}(t_{lips}; \frac{x_{eyes} \cdot x_{lips}}{modif_1})$$

$$S(t; x_{eyes}, x_{lips}) = \int_{t_{eyes}=t}^{\infty} \int_{t_{lips}=t}^{\infty} f_{eyes}(t_{eyes}; \frac{x_{eyes} \cdot x_{lips}}{modif_1}) \cdot S_{lips}(t_{eyes}; \frac{x_{eyes} \cdot x_{lips}}{modif_2}) \cdot f_{lips}(t_{lips}; \frac{x_{eyes} \cdot x_{lips}}{modif_2}) + \\ f_{lips}(t_{lips}; \frac{x_{eyes} \cdot x_{lips}}{modif_2}) S_{eyes}(t_{eyes}; \frac{x_{eyes} \cdot x_{lips}}{modif_1}) \cdot f_{eyes}(t_{lips}; \frac{x_{eyes} \cdot x_{lips}}{modif_1}) dt_{eyes} dt_{lips}$$

Note that direct non-selective influence is realized through multiplication of the parameters x_{lips}, x_{lips} between processing units. Thus, they both appear in each density. The modifiers of the rate parameters are denoted as $modif_1, modif_2$ and they are defined in a similar way to the cross-talk parameters in the dynamic systems model of Townsend and Wenger (2004).

As in the OR case, we will also assume following constraints:

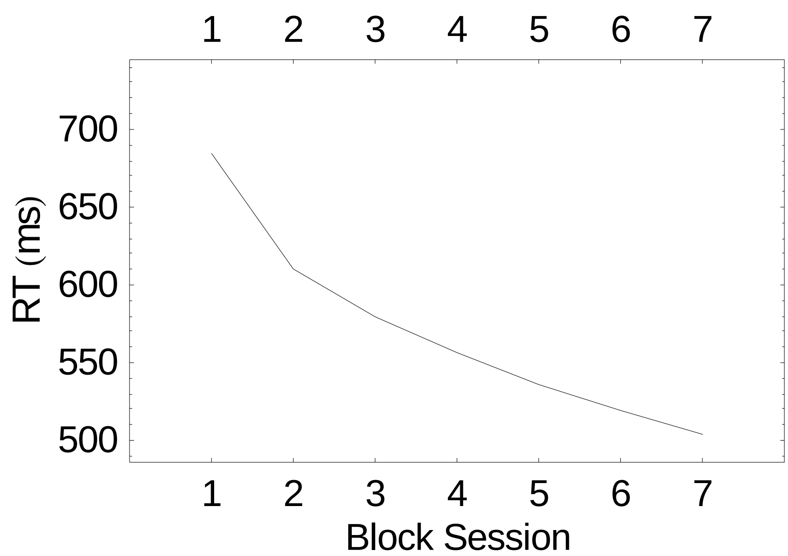
$$x_{lips}, x_{lips} = \{high, low\}$$
$$1 \leq x_{lips}, x_{lips} \leq 0$$
$$modif_1, modif_2 \leq 0$$
$$1 \leq high, low \leq 0$$
$$low \leq high$$

The model above can be reduced to an independent parallel exhaustive model if we set the values of the modifiers to $modif_1 = x_{lips}$ and $modif_2 = x_{eyes}$ in the above equation. The magnitudes of $\frac{1}{modif_1}, \frac{1}{modif_2}$ define the level of mutual interaction between processing units. As the value of that ratio is getting larger then the strength of positive dependency is larger as well.

SIC test implementation

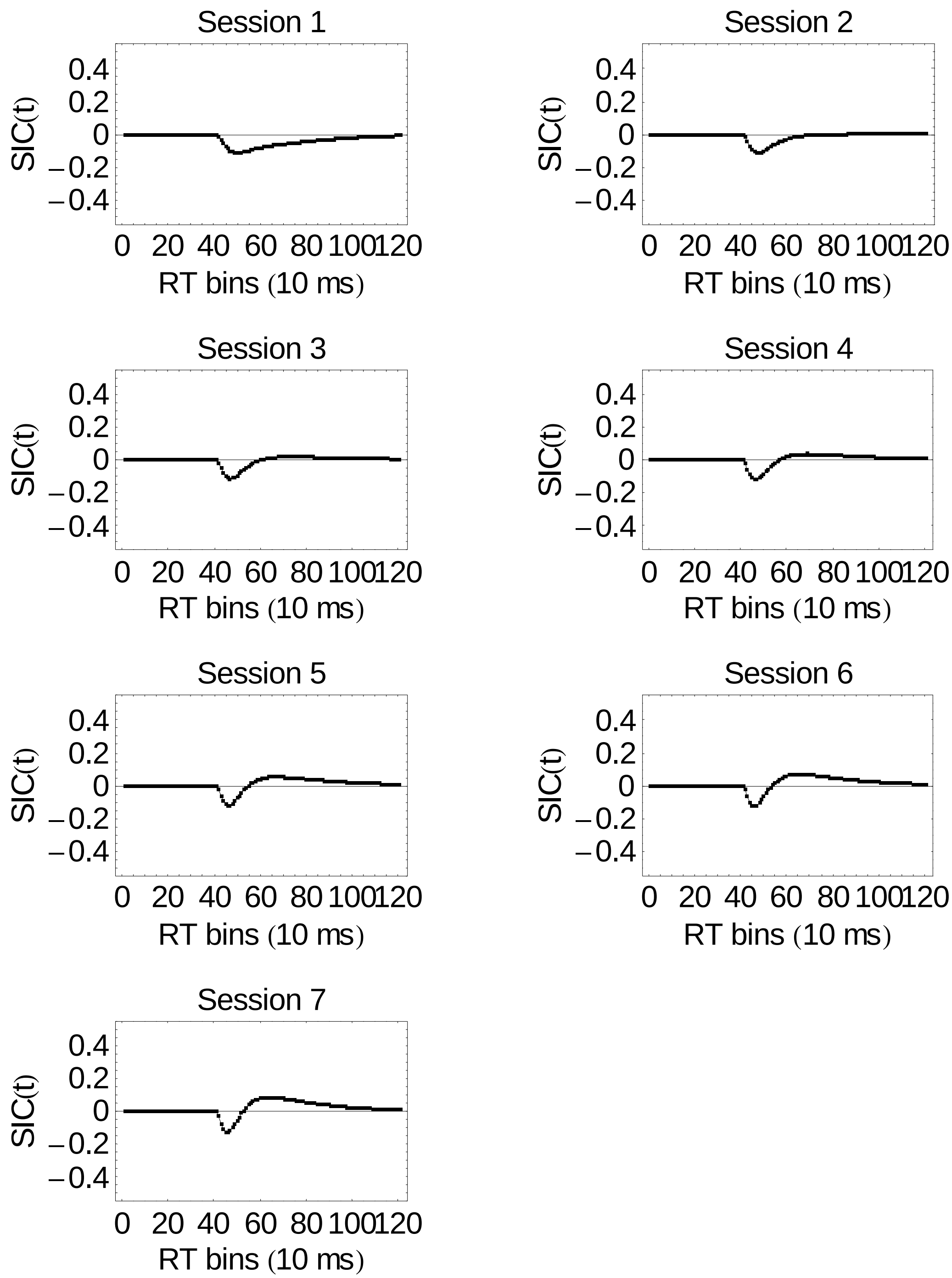
We will provide an example how both the SIC function and mean RT change as a function of learning. We assume that the rate parameters x_{lips}, x_{lips} are constant over learning, and that the modifiers will *monotonically* change their values over the course of learning. We also assume that the first session exhibits no between channel cross-talk, that is a parallel independent exhaustive architecture. Table 11 above lists values of parameters used in the simulation, where each row corresponds to one session (7 total). The values of parameters $modif_1, modif_2$ are calculated by the same function as in the OR condition.

Figure 55: Calculated mean RTs over learning sessions. The decreasing trend of RTs corresponds to the observed data on each group of faces in the AND experiments.



In Figure 56, we present the change of SIC functions over the different learning sessions. As we can see from Figure 56, the SIC shape starts as a primarily negative function, which is the signature of parallel exhaustive processing. Over the course of learning, the negative SIC function evolves into an S-shaped function that is predominantly positive by the final session.

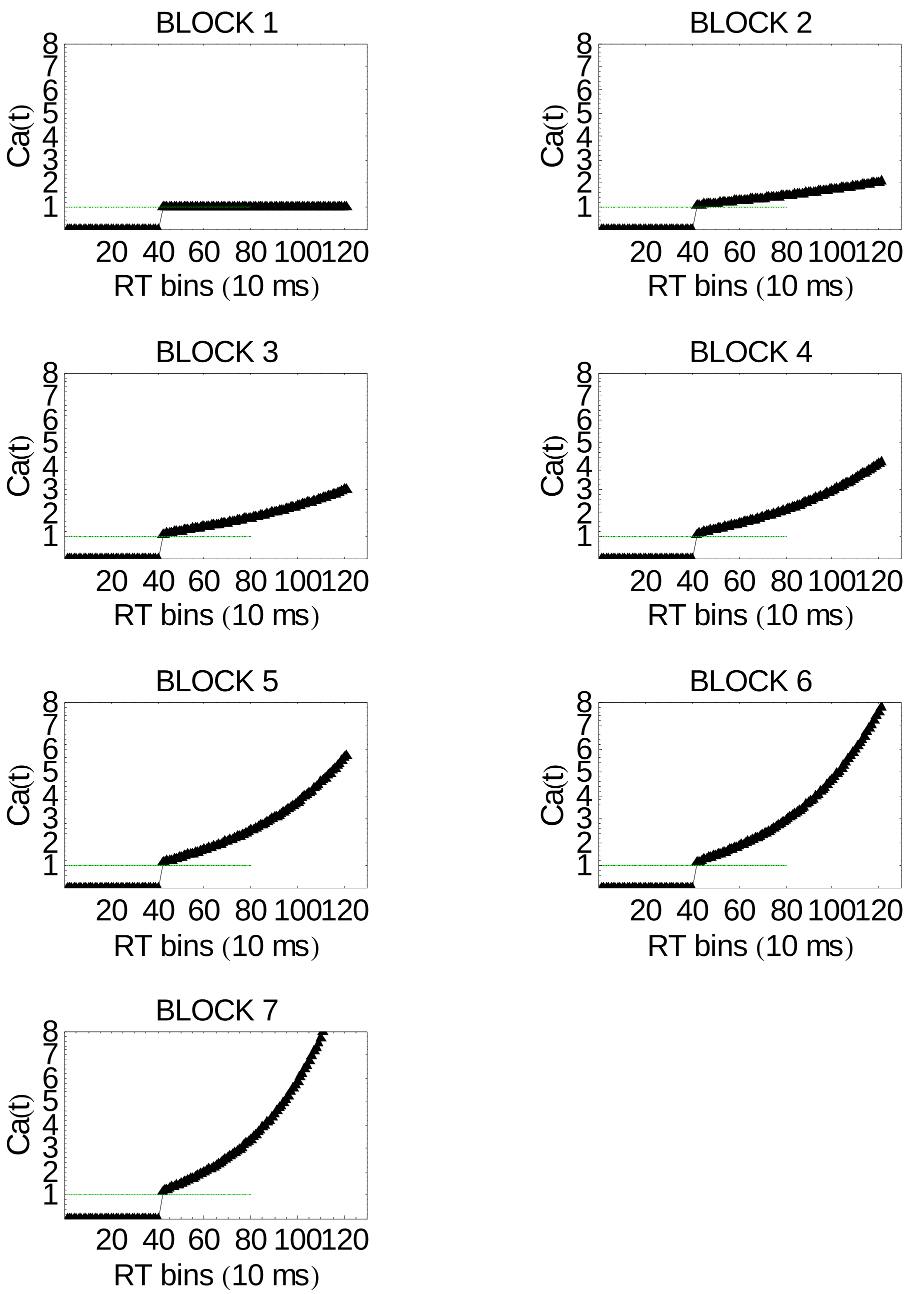
Figure 56: Simulated SIC functions for parallel positively dependent minimum-time (AND) processing model, across learning sessions.



Absolute Learning CCFs

In Figure 57, we present the absolute learning CCFs over different learning sessions. Recall that in the absolute learning CCF, the integrated hazard functions for the processing of parts are kept constant over the learning session. In order to characterize the processing of face parts (in the denominator) we arbitrarily set their respective values to $\{x_{lips}, x_{lips}\}/2$, for each face feature. As we can see from Figure 57 the magnitude of the CCF tends to increase with the increase in positive facilitation between units. This simulation provides, qualitatively, an account of our experimental findings.

Figure 57: Simulated the absolute capacity coefficient functions (CCFs) for parallel positively dependent minimum-time (AND) processing model, across learning sessions.



We conclude that mere manipulation of the cross-talk structure, realized through the modifying parameters, provided an appealing explanation for the data in the AND condition. All aspects - the mean RTs, overall shapes of SIC functions, and the CCFs - showed very similar qualitative patterns to the real observed trends as a function of learning. For example, observe the results that Participants 1 or 2 showed in the AND condition. By simple manipulation of the change of positive cross-talk between processing units, we are able to qualitatively account for the observed data.

Thus, we have demonstrated how the simplest possible manipulation of the system's structure can account for all the learning trends revealed in both the OR and AND conditions. This provides very strong support for the hypothesis regarding learning and holistic face effect outlined at the beginning of study. The direct implication is that a holistic face representation is an emerging property which relies mainly on the establishment of a dependency structure between feature representations.

Final conclusions

We conclude that the learning phase analysis and conclusions provided firm and unequivocal support for our approach by defining configurality in terms of characteristics of processing (SFT). More generally, gestalt/holistic/configural is *an emerging property,* as it could be learned. During the learning sessions, we observed the development of gestalt properties in the both AND and OR conditions at both the SIC and capacity levels.

The learning of faces is not based on a change of architecture of the system. Rather, a parallel architecture was preserved over all blocks of learning. What is learnt is a stochastic interdependence between two face features. We suggest that the dependence was based on positive facilitation between processing units, such that recognition of one face feature bolstered recognition of the other one. At this point, we can not decide whether this facilitation happened directly between the two face features or if it was indirect over some third agent. For example, it is possible that the dependency between the eyes and lips was realized through other background features, rather than directly between the two. The testing phase (second part of the experiment) will provide more information regarding the properties of this facilitation by comparison between the processing old faces used in the standard test and processing exhibited on new-faces that were either configurally of featurally manipulated.

The positive facilitation developed very early in learning, in the first two sessions. We can not provide an answer as to whether face processing is special given that we did not investigate a comparable case with a non-face object. It could be implied that face encoding relies on some very old, or even innate mechanisms, that gives a quick rise to super capacity in the first secessions. But more importantly, the emergence of an even more mutually facilitatory structure implies that learning can always play a significant role in face perception. Why, then, restrict the learning only to faces? It seems reasonable to assume that even a non-face object will show an increase of capacity as a function of learning. It is important to note that the strong or weak holistic hypothesis of object recognition should not depend on a starting level of learning during initial sessions, but rather whether the learned objects exhibit the same or similar trends of SIC and capacity (during learning and the configural test).

We also conclude that holistic/configural properties are not static. At the end of some learning period they do not necessarily stop and dig a trench at that point. The cognitive system could access different levels of a face representation on imaginary analytic/holistic dimension. What is demonstrated here is that sensitivity of whole face encoding depends on global experimental context: when whole faces are mixed with masked faces, recognition was slowed down, in contrast to when whole faces are presented in the blocked condition. We refer here to the blocked-mixed effect exhibited on both the mean RT level and capacity coefficient function level, which is stronger. When combined with part-based faces, whole-face encoding is slower. We suggest that the proposed processing system (Modeling section above) can be consciously controlled, at least at the level that regulates a level of mutual facilitation between learnt features. By that we mean that participants can adjust their global experimental strategies during face encoding such that they can attain more holistic/configural than analytic/featural properties. In our model, this is simply done by controlling the values of the mutual facilitators between channels. What was previously learned can be manipulated further. However, these issues are beyond the scope of this and more investigation is needed.

The test phase

In the test phase, we compared three experimental parts. In the first, which served as the control, participants perceived old faces. In the configural-test, we investigated the possibly detrimental change of face background to the recognition of faces. The old faces were replaced with the new face backgrounds, where only two important facial features (the eye-separation and lips-position) remained at their original spatial positions. This manipulation was analogous to the part-to-whole manipulation of Tanaka and Sengco (1997), which is considered to be the ultimate configural test. In the featural test, we completely removed the face background, and left only the two important face dimensions. This is the analogue of the recognition of a feature(s) in the isolation. The comparison of these test conditions will indicate the role of face background.

OR

In the standard test experiment, we replicated the findings from the learning session; that is, participants exhibited parallel self-terminating architecture with super capacity, revealed by the SIC and capacity tests. We confirmed previous findings obtained in the learning session. The configural test also exhibited parallel self-terminating architecture. Only 3 participants revealed super capacity, while the other half exhibited unlimited capacity. Additional tests based on statistical comparisons of bootstrapped confidence intervals around CCFs revealed that for all participants old faces exhibited a larger magnitude of capacity than in new-face condition (configurally altered). Moreover, in the feature-test condition, half of the participants exhibited parallel self-terminating architecture, while the other half showed non-regular MIC and SIC results, with the lips-position effect not significant. Their capacity functions were all of the smallest magnitude when compared with other conditions and never super capacity.

On the mean RT and mean accuracy levels we replicated the general findings necessary to establish the configural effect: old whole faces were processed faster and more accurately than the new faces and feature faces (criteria established after Tanaka & Farah, 1993; Tanaka & Sengco, 1997).

However, it was surprising to observe that new faces did not disrupt face recognition to a larger extent. In fact, architecture did not change with the configural manipulation. Instead the capacity has changed. We conclude that the configural effect defined as severely disrupting the face background produced a slight regression to the parallel self-terminating model, but still preserved facilitation between units.

More interestingly, the face features produced deeper regression by suppressing some perceptual effects. Participants' performance was not seriously affected, because in this task the feature detectors read from redundant sources and only one feature needed to be completely processed. The absence of the lips-position effect also makes sense because we removed the face background that cued the spatial position of the lips. When removed, the lips remained without local support. It is then evident that the learning did not establish a direct connection between the spatial positions of the eyes and lips. Rather, it seems that participants relied on local support from the surrounding face features not important for making decision.

Also, the difference between the standard and configural tests also supports the aforementioned idea that local background features aided recognition of the two critical features. This support comes from two sources. The first one is direct support for a connection established between, for example, the old nose and the eye-separation. When the old nose is replaced with the new nose, then that relationship disappeared. The second support comes through help in the first order spatial localization of the two important features. Face background has the simple role of providing tags use to globally locate the other features (first order relations). We propose that the new configuration forced participants to suppress the learnt relationship between any two face properties, in this example the nose and eye-separation. When the whole face background is removed, then both types of supports are lost, and participants could only use very local spatial codes for making the decision.

AND

In the standard test experiment, we replicated the findings from the learning session where participants exhibited mainly S-shaped SIC functions, which correspond to parallel exhaustive positively dependent architectures. Two participants exhibited mostly negative SIC functions which correspond to parallel exhaustive independent processing. Participants exhibited super capacity in almost all conditions, with different magnitudes. Additional tests based on statistical comparisons of bootstrapped confidence intervals around the CCFs revealed that for all participants, old faces exhibited larger magnitudes of capacity than in new-face conditions (configurally altered), with the exception of Participant 4. Moreover, in the feature-test condition, all participants exhibited parallel exhaustive architecture, with statistically regular MIC and SIC results. Their capacity functions were all of the smallest magnitude compared to the other conditions and were never super capacity.

On the mean RT and mean accuracy levels, we also replicated the general findings of the configural effect: old whole faces were processed faster and more accurately then the new faces and feature faces.

Similar to the OR condition, the configural disruption did not change the architecture. Instead, the capacity has changed. We conclude that the configural effect produced a moderate regression to the parallel exhaustive model, still preserving facilitation between the units, but to a lesser extent.

The face features produced a larger difference between old and feature faces. More interestingly, both face features (the eye-separation and lips-position) generated significant effects when the face background was removed. In contrast to the OR condition, the lips-position produced significant perceptual effects and obviously did not depend on the removed local features. In fact, we suggest that in the AND task, the integrative feature processes spread over larger face area on the face in contrast to the OR task. We suggest that the lips were stochastically correlated with the eyes during the learning, because a decision required the processing of both, and that remained during the featural-test, producing perceptual effects for both features. We also conclude that the role of the face background is important as the third agent because it helps the face features to establish the parallel positively dependent processing structure, which ultimately contributes to the some form of holistic representation.

Study of Wenger and Ingvalson (Wenger & Ingvalson 2002) showed not much support for the strong holistic hypothesis. In subsequent paper (Wenger & Ingvalson 2003) they replicated previous findings that revealed the decisional component was responsible for configural effects observed in their studies. They used the constructs of general recognition theory (GRT) in order to investigate the source of holism in face perception. Within the GRT framework, there are several instances at which holism could be produced. The *informational dependence* is defined as the dependency between processing (face) features and is realized within the single stimulus. This is considered as the strongest level of holism. The *informational separability* is defined as the ability to recognize one feature of stimulus (face) irrespective of changes of the level on the other feature. The *decisional separability* localizes the holism in the response criteria, by allowing the decision boundary for one feature to shift across levels of the other feature. The decisional separability is the weakest form of holism because it is not produced by the interaction between processing features, but by shift of a decisional response criteria. Direct consequence of the violation of the decisional separability could be that a face representation is encoded as a set of independent features, and not as a unified representation, as suggested by the strong holistic hypothesis. This approach provides information of representational status of holistic processing. Although the GRT provides more complex view into the system's structure, the issue of architecture and how processes are organized is not explained in the framework.

The natural question that arises is whether the results from our study could be explained by the failure of the decisional separability. The theory that can combined both the GRT and the SFT approach is not developed yet. But we can provided some reasoning here how the violation of the decisional separability can or can not be related to our findings. Note that we explained our results in terms of violations of the dependency between processing features (strong holism), but under the SFT approach.

All three constructs of the GRT mentioned above, the *informational dependence, informational separability*, *decisional separability* are logically independent of each other. Although it is possible that they interact in some peculiar way. At this point we could consider the case whey they are independent only (Appendix B). The proof of Proposition 2 demonstrates that even though the decisional separability could be violated in the SFT experimental design conditions, it would be violated in the same way in all factorial condition in the SFT design (HH, HL, LH and LL in 2 x 2 factorial design), and will not change the predictions for different architectures. If the decisional aspect is independent of what happens between two processing features, then the decisional aspect will not interfere with the SIC function appearance, qualitatively.

In the corollary of Proposition 2 we suggest that the findings regarding the importance of the decisional component in the holistic encoding (Wenger & Ingvalson 2003) can not be used as an explanation of exhibited behavior of change of SIC functions during learning sessions. Therefore, we suggest that the findings in our study support stronger holistic hypothesis, which assumes violation of independence between encoded face units. However, if there the decisional component is affected by how features are processed in term the speed and/or possible interaction between them, then no such conclusion can be accepted. We are looking forward the new unified theory of the GRT and SFT in order to provide stronger conclusion.

In this study, we aimed to investigate the processing characteristics of face recognition. We applied the stronger tests for the architecture and capacity of the system. We extended previous lines of investigation by controlling for different stopping rules – exhaustive and self-terminating - that were manipulated through the AND and OR conditions. In order to accommodate for typical part-to-whole paradigms that are considered to be the ultimate tests for holistic/configural encoding, we manipulated two face features in a factorial manner, and employed a categorization paradigm. Unlike previous studies (Tanaka & Farah, 1993; Tanaka & Sengco, 1997), we collected both RT and observed errors. The results replicated general findings using part-to-whole configural tasks on the error level (Tanaka & Sengco, 1997). The advantage of this study was demonstrated in the precise characterization of the processing properties of the face recognition system.

We propose the property of holism that is emergent and develops over time. It also depends on the nature of the task, and can be modulated to cover a large area of the face and produce large integrative effects of face units. Overall, processing exhibited a parallel architecture with positive cross-talk between features, that appear to be controlled by cognitive strategy or task expectations. The emergent property of holistic processing indicates that no static view of mental face representation or/and processing structure is plausible. So the real question is not what the **final** status of face representations is, for example. Or **exactly** how the structure is organized during face perception. The real issue is how the system changes with experience, in terms of both the representation and the mental organization, and what dynamical structure can accommodate for these changes.

We also propose that the cross-talk structure is closely related to what we observed as a configural face effect, and could be task-dependant. This also fits the view suggested under the strong holistic hypothesis, which assumed the interaction between featural and configural information. Our suggestion is that this interaction could be controlled, and that participants can exhibit cognitive control that range from analytic strategies, favoring independent features realized through no cross-talk, to extremely holistic processing strategy, such that they fuse all units into a single channel, which eventually seems to happen if we increase the cross-talk between units to a very large extent. It is also possible to exhibit in-between strategies, which could account for findings when participants favored features more then configurations, or vice versa.

We will conclude with the statement that this study represents an alternative theoretical view of the face recognition process. We questioned the validity of the theories based on attempts to operationally define the notion of holism, without systematic theoretical constructs. The strong tests coupled with simulations and stable patterns of data on the single participant's level, supported our research hypotheses.

References:

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*(1), 33-53.

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics, 6(6-A)*, 321-326.

Bartlett, J. C. (1993). Limits on losses in face recognition. In J. Cerella & J. M. Rybash (Eds.), *Adult information processing: Limits on loss.* (Vol. xxi, pp. 351-379). San Diego, CA: US: Academic Press,Inc.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*(3), 281-316.

Bartlett, J. C., Searcy, J. H., & Abdi, H. (2003). What are the routes to face recognition? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes.* (pp. 393). London: Oxford University Press.

Belhumeur, P. N., & Kriegman, D. J. (1997). *What is the set of images of an object under all possible lighting conditions?* Paper presented at the IEEE Conference on computer vision and pattern recongition.

Bentin, S., Allison, T., Puce, A., Perez, E., & et al. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*(6), 551-565.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103*(4), 751-757.

Bradshaw, J. L., & Wallace, G. (1971). Models for the processing and identification of faces.

Bruce, V., Dench, N., & Burton, M. (1993). Effects of distinctiveness, repetition and semantic priming on the recognition of face familiarity. *Canadian Journal of Experimental Psychology, 47*(1), 38-60.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38*(2), 109-144.

Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. In V. a. H. Bruce, Glyn W (Ed.), *Special issue of Visual Cognition* (Vol. Vol. 1, pp. 141-180): Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Busey, T. A. (2001). Formal models of familiarity and memorability in face recognition. In Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Scientific psychology series* (pp. 147-191). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45*, 431-448.

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. *Memory & Cognition, 27*(1), 139-151.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science, 11*(5), 429-433.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology, 34*(3), 253-275.

Colonius, H., & Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 245-254). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Craw, I., & Cameron, P. (1996). *Face recognition by computer.* Paper presented at the In Proceedings, British Machine Vision Conference.

Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *American Journal of Psychology. Special Views and varieties of automaticity, 105*(2), 271-315.

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica, 73*(3), 225-243.

Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the Photofit technique. *Perception, 6*(3), 263-269.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision* (No. 14 ed. Vol. viii). Oxford psychology series London: Oxford University Press.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 39*(2), 197-215.

Donnelly, N., & Davidoff, J. (1999). The mental representations of faces and houses: Issues concerning parts and wholes. *Visual Cognition. Special Issue: The neuroscience of perceptual integration, 6*(3-4), 319-343.

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychonomic Bulletin & Review, 6*(2), 239-268.

Edelman, S., & O'Toole, A. J. (2001). Viewpoint generalization in face recognition: The role of category-specific processes. In. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, London: Chapman & Hall.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). Emotion in the human face: Guidelines for research and an integration of findings.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology, 30*, 527-554.

Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology, 66*(4), 409-426.

Enns, J. T., & Gilani, A. B. (1988). Three-dimensionality and discriminability in the object-superiority effect. *Perception & Psychophysics, 44*(3), 243-256.

Farah, M., Levinson, K. L., & Klein, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia, 33*(6), 661-674.

Farah, M. J.,(1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science, 1*(5), 164-169.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology. Special The cognitive neuroscience of face processing, 17*(1-3), 117-123.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception & Performance, 21*(3), 628-634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482-498.

Galton, F. (1879). Composite portraits made by combining those of many different persons into a single resultant figure. *Nature, 18*, 97-100.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*(1), 98-123.

Gauthier, I., & Curby, K. M. (2005). A Perceptual Traffic Jam on Highway N170: Interference Between Face and Car Expertise. *Current Directions in Psychological Science, 14*(1), 30-33.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191-197.

Gauthier, I., & Tarr, M. J., (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance, 28*(2), 431-446.

Ginsburg, A. P. (1978). *Visual information processing based on spatial filters constrained by biological data.* Unpublished Doctoral dissertation, Cambridge University, England.

Graham, N. V. S. (1989). *Visual pattern analyzers* (Vol. xvi). New York, NY: Oxford University Press.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics, 35*(5), 451-463.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*, 555-562.

Groner, R. (1967). Perceptual Dimensions of Faces. *Zeitschrift für Experimentelle und Angewandte Psychologie, 14*(1), 135-154.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception, 13*(5), 505-512.

Haig, N. D. (1985). How faces differ--a new comparative technique. *Perception, 14*(5), 601-615.

Harmon, L. D. (1973). The recognition of faces. *Scientific American, 229*(5), 71-82.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23*(1), 65-74.

Hole, G. J., George, P. A., & Dunsmore, V. (1999). Evidence for holistic processing of faces viewed as photographic negatives. *Perception, 28*(3), 341-359.

Homa, D., Haver, B., & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition, 4*(2), 176-185.

Hosie, J. A., Ellis, H. D., & Haig, N. D. (1988). The effect of feature displacement on the perception of well-known faces. *Perception, 17*(4), 461-474.

Huang, J., Heisele, B., & Blanz, V. (2003). *Component-based face recognition with 3D morphable models.* Paper presented at the International Conference on Audio- and Video-Based Person Authentication.

Ingvalson, E. M., & Wenger, (2005). A strong test of the dual-mode hypothesis. *Perception & Psychophysics, 67*(1), 14-35.

Inn, D., Walden, K. J., & Solso, R. L. (1993). Facial prototype formation in children. *Bulletin of the Psychonomic Society, 31*(3), 197-200.

Innes-Ker, A. H. K. (2003). *Gestalt perception of emotinal expressions.* Unpublished Ph.D., Indiana University, Bloomington.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*(1-2), 1-19.

Kanade, T. (1973). *Computer recognition of human faces*. Birkhauser, Basel, Switzerland, and Stuttgart, Germany.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*(1), B1-B11.

Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12.*

Koffka, K. (1935). Principles of Gestalt psychology.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51A*(3), 449-473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A*(2), 513-536.

Logan, G. D., & Stadler, M. A. (1991). Mechanisms of performance improvement in consistent mapping memory search: Automaticity or strategy shift? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*(3), 478-496.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology: I* (pp. 491). Oxford, England: John Wiley.

Macho, S., & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *Journal of Experimental Psychology: Human Perception & Performance, 24*(5), 1486-1500.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition. Scientific psychology series* (pp. 147-180). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Matthews, M. L. (1978). Discrimination of Identikit constructions of faces: Evidence for a dual processing strategy. *Perception & Psychophysics, 23*(2), 153-161.

McKelvie, S. J. (1973). The meaningfulness and meaning of schematic faces. *Perception & Psychophysics, 14*(2), 343-348.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207-238.

Mermelstein, R., Banks, W. P., & Prinzmetal, W. (1979). Figural goodness effects in perception and memory. *Perception & Psychophysics, 26*(6), 472-480.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247-279.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception & Performance, 17*(1), 160-169.

Milord, J. T. (1978). Aesthetic aspects of faces: A (somewhat) phenomenological analysis using multidimensional scaling methods. *Journal of Personality & Social Psychology, 36*(2), 205-216.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*(5), 553-566.

Mordkoff, J. T., & Egeth, H. E. (1993). Response time and accuracy revisited: Converging support for the interactive race model. *Journal of Experimental Psychology: Human Perception & Performance, 19*(5), 981-991.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception & Performance, 17*(2), 520-538.

Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence of coactivation. *Perception & Psychophysics, 53*(4), 357-366.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*(5), 555-604.

Nefian, A. V., & Hayes III, M. H. (1998). *Hidden Markov models for face recognition.* Paper presented at the International Conference on Acoustics,

Speech and Signal Processing.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13*(1), 87-108.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance, 17*(1), 3-27.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*(2), 266-300.

Okada, K., Steffans, J., Maurer, T., Hong, H., Elagin, E., Neven, H., et al. (1998). The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test. In P. J. P. H. Wechsler, V. Bruce, F. F. Soulie, and T. S. Huang (Ed.), *Face Recognition: From Theory to Applications,* (pp. 186–205). Berlin, Germany: Springer-Verlag.

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

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*(4), 291-303.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning & Memory, 3*(4), 406-417.

Pentland, A., Moghaddam, B., & Starner, T. (1994). *View-based and modular eigenspaces for face recognition.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition.

Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology, 93*(1), 1-30.

Rakover, S. S., & Teucher, B. (1997). Facial inversion effects: Parts and whole relationship. *Perception & Psychophysics, 59*(5), 752-761.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275-280.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*(1), 161-168.

Rhodes, G. (1985). Lateralized processes in face recognition. *British Journal of Psychology, 76*(2), 249-271.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17*(1), 43-63.

Rhodes, G. (1993). Configural coding, expertise, and the right hemisphere advantage for face recognition. *Brain & Cognition, 22*(1), 19-41.

Rossion, B., Dricot, L., Devolder, A., Bodart, J.-M., Crommelinck, M., de Gelder, B., et al. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 12*(5), 793-802.

Schweickert, R. (1978). A critical path generalization of the additive factor methods analysis of a stroop task. *Journal of Mathematical Psychology, 26*, 105-139.

Schweickert, R. (1983). Latency network theory: Scheduling of processes in sentence verification and the Stroop effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 9*(3), 353-383.

Schweickert, R., Giorgini, M., & Dzhafarov, E. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *Journal of Mathematical Psychology, 44*(4), 504-535.

Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology, 33*(3), 328-347.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception & Performance, 22*(4), 904-915.

Sergent, J. (1982). About face: Left-hemisphere involvement in processing physiognomies. *Journal of Experimental Psychology: Human Perception & Performance, 8*(1), 1-14.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75*(2), 221-242.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*(4), 373-421.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*(4468), 390-398.

Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General, 115*(1), 58-61.

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 53-71). Washington, DC: American Psychological Association.

Shepard, R. N., & Chang, J.-J. (1963). Stimulus generalization in the learning of classifications.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). *Learning and memorization of classifications*.Unpublished manuscript.

Shepherd, J. W., Davies, G., & Ellis, H. D. (1981). Studies of cue saliency. In G.Davies, H. Ellis & J. Shepherd (Eds.), *Perceiving and remembering faces*. London: Academic Press.

Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces--a comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20*(2), 125-133.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127-190.

Smith, E. E., & Medin, D. L. (2002). The exemplar view. In D. J. Levitin (Ed.), *Foundations of cognitive psychology: Core readings.* (pp. 277-292). Cambridge, MA, US: MIT Press.

Smith, E. E., & Nielsen, G. D. (1970). Representations and retrieval processes in short-term memory: Recognition and recall of faces.

Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology, 72*(4), 499-503.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153(3736)*, 652-654.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, Amsterdam, 30*, 276-315.

Steyvers, M., & Busey, T. A. (2001). Predicting similarity ratings to faces using physical descriptions. In Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Scientific psychology series* (pp. 115-145). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Swets, D. L., & Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18,*, 831–836.

Tanaka, J. W., & Farah, (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics, 50*(4), 367-372.

Tanaka, J. W., & Farah, M. J., (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A*(2), 225-245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*(5), 583-592.

Tarr, M. J., (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2*(1), 55-82.

Tarr, M. J.,(2003). Visual object recognition: Can a single mechanism suffice? In M. A. R. Peterson, Gillian (Ed.), *Perception of faces, objects, and scenes: Analytic and holistic processes. Advances in visual cognition*. London: Oxford University Press.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences, 7*(1), 23-30.

Thomas, R. D. (2001a). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Scientific psychology series* (pp. 193-227). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Thomas, R. D. (2001b). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics, 63*(4), 625-650.

Thorpe, S., & Fabre-Thorpe, M. l. (2001). Seeking categories in the brain. *Science, 12*, 260-263.

Townsend, J. T. (1971a). A note on the identifiability of parallel and serial processes. *Perception and Psychophysics, 10*, 161-163.

Townsend, J. T. (1971b). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics, 10*, 161-163.

Townsend, J. T. (1972). Some results concerning the indentifiability of parallel and serial processes. *British Journal of Mathematical & Statistical Psychology, 25*(2), 168-199.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28*(4), 363-400.

Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin, 108*(3), 551-567.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive Theory* (Vol. III, pp. 200-239): Hillsdale, NJ: Erlbaum Associates.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*: Cambridge, UK, Cambridge Univ. Press.

Townsend, J. T., & Colonius, H. (1997). Parallel processing response times and experimental determination of the stopping rule. *Journal of Mathematical Psychology, 41*(4), 392-397.

Townsend, J. T., & Landon, D. E. (1982). An experimental and theoretical investigation of the constant-ratio rule and other models of visual letter confusion. *Journal of Mathematical Psychology, 25*(2), 119-162.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*(4), 321-359.

Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology, 33*(3), 309-327.

Townsend, J. T., Solomon, B., & Smith, J. S. (2001). The perfect Gestalt: Infinite dimensional Riemannian face spaces and other aspects of face perception. In Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Scientific psychology series* (pp. 39-82). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Townsend, J. T., & Spencer-Smith, J. (2004). Two Kinds of Global Perceptual Separability and Curvature. In C. Kaernbach & E. Schroger (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition.* (pp. 89-109). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Townsend, J. T., & Thomas, R. D. (1993). On the need for a general quantitative theory of pattern similarity. In S. C. Masin (Ed.), *Foundations of perceptual theory.*

Advances in psychology, Vol. 99 (pp. 297-368). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Townsend, J. T., & Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology, 38*(1), 1-34.

Townsend, J. T., & Wenger, (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*(3), 391-418.

Townsend, J. T., & Wenger, (2004). A Theory of Interactive Parallel Processing: New Capacity Measures and Predictions for a Response Time Inequality Series. *Psychological Review, 111*(4), 1003-1035.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience. Special Face perception, 3*(1), 71-86.

Tversky, A., & Krantz, D. H. (1969). Similarity of schematic faces: A test of interdimensional additivity.

Ullman, S. a. B., R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*(10), 992-1006.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43A*(2), 161-204.

Van Zandt, T., & Townsend, J. T. (1993). Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception & Psychophysics, 53*(5), 563-580.

Walker-Smith, G. J. (1978). The effects of delay and exposure duration in a face recognition task. *Perception & Psychophysics, 24*(1), 63-70.

Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object-superiority effect. *Science, 186*(4165), 752-755.

Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 872-892

Wenger, M. J., & Ingvalson, E. M. (2003). Preserving Informational Separability and Violating Decisional Separability in Facial Perception and Recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*(6), 1106-1118.

Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *Journal of General Psychology, 127*(1), 67-99.

Wenger, M. J. & Townsend, J. T. (2001). Faces as Gestalt stimuli: Process characteristics. In Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. Scientific psychology series* (pp. 229-284). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Wheeler, D. D. (1970). *Processes in the visual recognition of words.* Unpublished Dissertation, U. Michigan.

Winograd, E. (1976). Recognition memory for faces following nine different judgments. *Bulletin of the Psychonomic Society, 8*(6), 419-421.

Wiskott, L., Fellous, J. M., Kruger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*, 775-779.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141-145.

Young, A. W., & De Haan, E. H. F. (1995). Face recognition and awareness after brain injury. In A. M. M. Rugg (Ed.), *The neuropsychology of consciousness*. New York: Academic Press.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.

Yuille, A. L. (1991). Deformable templates for face recognition. *Journal of Cognitive Neuroscience. Special Face perception, 3*(1), 59-70.

Zandt, T. V. (2002). Analysis of response time distributions. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology. Methodology in experimental psychology.* (3rd ed., pp. 461-516). New York, NY: John Wiley & Sons, Inc.

Zhao, W., Chellappa, R., & Krishnaswamy, A. (1998). *Discriminant analysis of principal components for face recognition.* Paper presented at the In Proceedings, International Conference on Automatic Face and Gesture Recognition.

Appendix A

The capacity bound for the masked whole-faces

Some concerns have been raised about using the masked faces in order to calculate the capacity coefficient function in both the OR and AND conditions. Recall that for the OR condition (self-terminating), the capacity coefficient function is defined as:

$$C_0(t) = \frac{H_{AB}(t)}{H_A(t) + H_B(t)}$$

Usually, both denominator terms are calculated from processing of a single processing feature (Townsend & Nozawa, 1995). We will use a face study example: the H_{AB} stands for a whole face that consists of two important features, while H_A and H_B stand for the single face features, presented on separate trials. In this study we used the masked whole faces in order to calculate the denominator values (see method section). Both denominator conditions are designed by applying a shading mask over a whole face: two complementary masks based on pixel brightness level were produced (Figure 11). The concern was that this manipulation could change the capacity bound of C(t)=1, which is the bound of the unlimited capacity independent parallel processing system (UCIP). It is possible that the masks will allow for some non-linear effects in face perception such that the complementary masked part provides more information than a single face feature.

In order to show why our approach of using the masked faces is more beneficial, we will emphasize two aspects: first, the importance of using the whole-masked faces rather than feature-based faces for denominators; second, we will provide more experimental evidence testing capacity bounds of the masked whole faces and the single features.

Several reasons motivated the utilization of the masked whole faces in the capacity. The main goal of this study was to minimize the use of featural sources of face information. Our concern is that we should emphasize holistic strategies in face recognition, rather than analytic ones where participants could focus most of their attention in learning what the parts are (so called part-whole paradigm). The masked whole faces appear like whole faces in a more realistic, everyday background, like a face that you can see under the shade of a tree, chandelier or curtain. Additionally for the AND condition, a single feature trial can not be incorporated into the original task. Note that all Sharks faces in the AND condition share both their critical features with some Jets faces. Feature sharing design led to successful application of an the exhaustive processing rule with the Sharks, which was necessary for the study.

In order to investigate possible bound changes when the masked whole faces were used, we ran an additional experiment in which we wanted to compare mean processing times for masked and single-feature faces.

Method

Participants

Six new participants were paid for participation.

Materials

In the OR experiment we used the HH face from the gang-member group, and the hero face. We used the most distinctive faces because of the short duration of this experiment. Three types of faces were used for each of the gang member and hero faces: whole faces, masked faces and half faces (see Figure 58).

Figure 58: A whole face, its complementary masked faces, and its half-faces.

Design and procedure

Each participant accomplished two 45 minutes sessions, each run on separate days.

Each session started with 100 practice trials that consisted only of gang-member and hero whole face presentations. The hero face was presented on half of the trials, and the gang-member on another half trials. A participant had to decide, by pressing one of the mouse keys with the left and right index fingers, whether the hero or the gang-member was presented. RT was recorded from the onset of stimulus display, up to the time of response. Each trial consisted of a central fixation point (crosshair) for 1070ms followed by a high-pitch warning tone which lasted for 700msec. After that, a face was presented for 190ms. Upon incorrect categorization participants received an error message.

Each face manipulation for both the hero and gang faces was presented on 240 trials in total, equally dived by the hero and gang-member.

Results

We run repeated measures GLM on the six subjects. The main effect of face manipulation was significant ($F(1.083, 5.417)=61.09$, $p<0.001$, power=1), corrected by Greenhouse-Geisser criterion, because the assumption of sphericity was significantly violated. Mean RT for each face type is presented in Table 12. It can be observed that on average the masked faces are recognized faster than half-faces.

Table 12: Mean RTs for each face type, along with mean standard errors.

	Mean RT (ms)	StErrorMean
Whole-faces	567	25
Masked-faces	634	20
Half-faces	672	28

We ran pretest contrast analysis between the mean RT from Table 12, using the Bonferroni adjustment for multiple comparisons. The contrast of interest between the masked faces and half faces was not statistically significant, with $p=0.103$. However, it was significant using the LSD contrast test, $p<0.05$. We conclude that there is a possibility that using the masked faces can produce a change of the capacity bound $C_o(t)=1$.

The masked faces yield faster recognition than the half faces, and that can influence our inference of the capacity exhibited in different situations. Since the difference between the two types of stimuli favored faster processing of masked faces we will investigate the consequences of a possible capacity bound movement. In fact, it is possible to establish the following mathematical proof:

Proposition 1: If we allow some process to speed-up one or both of the two single-features trial conditions, which are used to calculate the two integrated hazard functions in the denominator, then the unlimited capacity bound for minimum time (OR) independent processing will be smaller than $C_o(t)=1$.

Proof:

The capacity function is defined as

$$C_o(t) = \frac{H_{a,b}(t)}{H_a(t) + H_b(t)} = \frac{-\ln(S_{a,b}(t_{a,b}))}{-\ln(S_a(t)) - \ln(S_b(t))} = \frac{\ln(S_{a,b}(t_{a,b}))}{\ln(S_a(t)) + \ln(S_b(t))}$$

Where $S_{a,b}(t_{a,b})$ is the joint survivor function for two simultaneously processed units, and $S_a(t)$, $S_b(t)$ are two survivor functions that each corresponds to a single processed unit. When minimum time parallel independent architecture is employed then $C_o(t)$=1 (Townsend & Nozawa, 1995).

Let's assume another function

$$C_o'(t) = \frac{H_{a,b}(t)}{H_{a'}(t) + H_{b'}(t)} = \frac{\ln(S_{a,b}(t_{a,b}))}{\ln(S_{a'}(t)) + \ln(S_{b'}(t))}$$

that corresponds to another experiment. Let us assume that in this experiment, single units are affected such that following hold:

$S_a(t) > S_{a'}(t)$ and/or $S_b(t) > S_{b'}(t)$, $t \in \{0,\infty\}$

such that for all,

$E[t_a] > E[t_{a'}]$ and/or $E[t_b] > E[t_{b'}]$.

Then it follows that after applying the natural logarithm, the following holds, given that it is a monotonic and well-behaved transformation we can assume

$|\ln(S_a(t))| < |\ln(S_{a'}(t))|$ and/or $|\ln(S_b(t))| < |\ln(S_{b'}(t))|$

which directly implies,

$\ln(S_a(t)) + \ln(S_b(t)) < \ln(S_{a'}(t)) + \ln(S_{b'}(t))$

If the numerator term $\ln(S_{a,b}(t_{a,b}))$ is common for both functions it directly follows that,

$C_o(t) > C_o'(t)$.

Thus, even the though processing time between the *a'* and *b'* units could be independent, the bound for unlimited capacity exhibited in the $H_{a,b}$ condition will be lower than $C_o(t)=1$. If one adopts the standard UCIP bound at $C_o(t)=1$ to infer capacity level (super, unlimited or limited), this bound will be more conservative, than when processes *a* and *b* were used. Thus the upper violation above the standard bound will indicate super capacity. In other words, in the case of the *a'* and *b'* units, the system has to exhibit more super capacity than in the $H_{a,b}$ condition in order to violate $C_o(t)=1$.

End of proof □

Corollary of Proposition 1: If two parallel minimum time processes a' and b' can exhibit positive facilitatory interdependence in each trial condition (for both terms the denominator) then such processing will reduce the bound for detecting super capacity.

Similar to Proposition 1 if two processes exhibit positive dependence then the same mathematical reasoning will follow, provided that the numerator term exhibits similar behavior. In short, any violation of C(t)=1 toward super capacity will indicate super capacity at an even stronger level, because the system has to exhibit more positive dependency in the numerator term to overcome the denominator terms.

An example could be the complementary masks used in our study. If it is possible that the single features of interest (either eye-separation or lips-position), can exhibit additional positive interaction in each masked face, used as one of the denominator terms, then processing time will speed-up for that component. Consequently, the whole face (in the numerator) has to exhibit even more positive dependency in order for the whole function to be $C_o(t)>1$, and super capacity.

Given that in the additional experiment we demonstrated that masked faces can result in faster processing than the half faces, we conclude that by adopting $C_o(t)=1$ as the UCIP bound, we make the capacity test more conservative.

Similar logic will apply for the AND condition, because of the symmetry of equations. However, we have not explicitly proved this case, given that this task involving the half faces, masked and whole faces in the AND condition ought to be more complicated. Caution is raised because the half faces do note require the exhaustive rule condition, while the masked- and whole faces are part of the AND condition (exhaustive strategy), and can not be combined in the same task.

We adopt the more conservative UCIP bound C(t)=1. We also rely on the weak hypothesis of the presence of a super capacity: CCFs should show an increasing trend in their magnitudes as a function of learning, at least. Recall that the presence of holistic properties is realized through the positive mutual facilitation between units, and should be revealed by an increasing trend of CCF magnitude over the learning sessions.

Appendix B

Wenger and Ingvalson (Wenger & Ingvalson 2002; Wenger & Ingvalson 2003) used the general recognition theory (GRT) in order to assess characteristics of face recognition. They showed that the decisional component was responsible for configural effects observed in face-perception, which supported weak holistic hypothesis. Since no exact theory exist how the GRT and SFT are related, it is not clear whether the findings in this study could be explained by the failure of decisional component (as defined in the GRT). Here we provide the mathematical proof for the proposition which states that if the decisional component and stochastic dependency are independent, in their nature, then failure of decisional component will not affect the outcome of the SIC calculation on qualitative level.

Proposition 2: adding random variable to some processing architecture (serial or parallel) that is calculated by the survivor interaction contrast function (SIC) will not qualitatively a shape of the SIC function (a sign a function and a number of crossings of x-axis), if the added r.v. is independent of processing properties in all factorial conditions of the SIC.

Proof:

Serial exhaustive case

The survivor interaction contrast could be defined as convolution of products of the difference between marginal density and survivor functions:

$$SIC(t_2)=\int_{t1=0}^{t2}\left(f_{x2}(t_1)-f_{x1}(t_1)\right)\cdot\left(S_{y2}(t_2-t_1)-S_{y1}(t_2-t_1)\right)dt_1$$

In order to add the base time r.v. we will convolute its density and distribute the terms:

$$SIC(t_3)=\int_{t2=0}^{t3}\int_{t1=0}^{t2}\left(f_{x2}(t_1)-f_{x1}(t_1)\right)\cdot\left(S_{y2}(t_2-t_1)-S_{y1}(t_2-t_1)\right)dt_1\cdot f_{base}(t_3-t_2)\;dt_2$$

$$SIC(t_3)=\int_{t2=0}^{t3}\int_{t1=0}^{t2}\left(f_{x2}(t_1)-f_{x1}(t_1)\right)\cdot\left(S_{y2}(t_2-t_1)-S_{y1}(t_2-t_1)\right)\cdot f_{base}(t_3-t_2)\;dt_1dt_2$$

$$SIC(t_3)=\int_{t2=0}^{t3}\int_{t1=0}^{t2}\left(f_{x2}(t_1)\cdot f_{base}(t_3-t_2)-f_{x1}(t_1)\cdot f_{base}(t_3-t_2)\right)\cdot\left(S_{y2}(t_2-t_1)-S_{y1}(t_2-t_1)\right)\;dt_1dt_2$$

The convolution allows for rearranging the terms in the equation

$$SIC(t_3)=\int_{t2=0}^{t3}\int_{t1=0}^{t2}\left(f_{x2}(t_1)\cdot f_{base}(t_2-t_1)-f_{x1}(t_1)\cdot f_{base}(t_2-t_1)\right)\cdot\left(S_{y2}(t_3-t_2)-S_{y1}(t_3-t_2)\right)\;dt_1dt_2$$

$$SIC(t_3)=\int_{t2=0}^{t3}\int_{t1=0}^{t2}\left(f_{x2}(t_1)\cdot f_{base}(t_2-t_1)-f_{x1}(t_1)\cdot f_{base}(t_2-t_1)\right)\cdot dt_1\left(S_{y2}(t_3-t_2)-S_{y1}(t_3-t_2)\right)\;dt_2$$

and since the convolution of two density function is another density function

$$SIC(t_3)=\int_{t2=0}^{t3}\left(f_{x2+base}(t_2)-f_{x1+base}(t_2)\right)\cdot\left(S_{y2}(t_3-t_2)-S_{y1}(t_3-t_2)\right)\;dt_2$$

If rename variables the outcome is formally equivalent to the first expression above,

$$SIC(t_3)=\int_{t2=0}^{t3}\left(f_{x2'}(t_2)-f_{x1'}(t_2)\right)\cdot\left(S_{y2}(t_3-t_2)-S_{y1}(t_3-t_2)\right)\;dt_2$$

Provided that x2’>x1’ and y2>y1, and consequently $S_{x2}(t)<S_{x1}(t)$and $S_{y2}(t)<S_{y1}(t)$, for all *t*.

The necessity argument can not be provided given that it is possible that some dependent serial exhaustive system can produce overall additivity, and can produce the same outcome as the independent and selectively influenced system.

For the parallel architecture we use similar argumentation. Let assume parallel independent architecture with two processes, then factorial combination in terms of the SIC function could be presented as:

Parallel exhaustive processing

$$SIC(t) = -\left(F_{x2}(t) - F_{x1}(t)\right) \cdot \left(F_{y2}(t) - F_{y1}(t)\right)$$

Given that $F_{x2}(t) > F_{x1}(t),\ F_{y2}(t) > F_{y1}(t)$ for all t={0,∞}, and x2>x1 and y2>y1

Adding a base time r.v. by convolution

$$SIC(t) = \int_{t1=0}^{t} -\left(F_{x2}(t) - F_{x1}(t)\right) \cdot \left(F_{y2}(t) - F_{y1}(t)\right) \cdot f_{base}(t - t_1) dt_1$$

$$SIC(t) = \int_{t1=0}^{t} -\left(F_{x2}(t) - F_{x1}(t)\right) \cdot \left(F_{y2}(t) \cdot f_{base}(t - t_1) - F_{y1}(t) \cdot f_{base}(t - t_1)\right) dt_1$$

$$SIC(t) = \int_{t1=0}^{t} -\left(F_{x2}(t) - F_{x1}(t)\right) \cdot \left(F_{y2+base}(t) - F_{y1+base}(t)\right) dt_1$$

We can rename the variables:

$$SIC(t) = \int_{t1=0}^{t} -\left(F_{x2}(t) - F_{x1}(t)\right) \cdot \left(F_{y2'}(t) - F_{y1'}(t)\right) dt_1$$

given that

$F_{y2'}(t) > F_{y1'}(t)$, for all *t.*

So we showed that the SIC value is not affected by adding the r.v. base time which is identical for all factorial conditions. Similarly we derive the proof for parallel exhaustive processing, and serial minimum time processing. Again, the necessity argument is can not be provided from the same reasons from above.

End of proof □

Corollary of Proposition 2: If the decision time r.v. is independent of both the base time and the architecture processing time (for both serial and parallel architectures), and is identically distributed across different factorial conditions in the SIC function, then its inclusion will not qualitatively (a sign a function and a number of crossings of x-axis) affect the shape of the SIC function.

Curriculum Vitae
Mario Fific

Mario Fific, Ph.D.
Cognitive Psychology, Cognitive Science
Indiana University
1101 E. 10th St.
Bloomington, IN 47405-7007
E-mail: mfific@indiana.edu
Phone: 812-855-1554

Education

2005 – Post doctorate with Robert Nosofsky, Indiana University.

2005 - Joint Ph.D. in Cognitive Psychology and Cognitive Science, with Certificate in Mathematical Modeling

1999 - A member of James Townsend's laboratory;
Doctoral candidate for joint Ph.D. in Cognitive Psychology and Cognitive Science, with Certificate in Mathematical Modeling,
Indiana University; Bloomington, Indiana

1994 -   MA program in Psychology
Finished all relevant courses, did not defend MA work
Later published in LEP reports (1999)
Department of Psychology
Faculty of Philosophy
University of Belgrade, Serbia and Montenegro

1991 - 1994 BA in Psychology
Department of Psychology
Faculty of Philosophy
University of Belgrade, Serbia and Montenegro

Awards and fellowships

2000 Indiana University Cognitive Science Program Summer research Fellowship
2001 Indiana University Cognitive Science Program Summer research Fellowship
2002 Indiana University Cognitive Science Program Summer research Fellowship
2003 Indiana University Cognitive Science Program Summer research Fellowship

2004 Indiana University Cognitive Science Program Summer research Fellowship

Employment, teaching experience and assistantships

1996 - 1999	A member of technical staff on project of Corpus of Serbian Language Laboratory for Experimental Psychology, Faculty of Philosophy and Institute for Experimental Phonetics and Speech Pathology, Belgrade, Serbia and Montenegro

Instructor

Experimental research methods, Indiana University (2002, fall)

Teaching Assistant:

1992	Tutor, summer session, in Petnica Science Center (center for advanced science and technology education of gifted children) Petnica, Serbia and Montenegro
1994 - 1999	Teaching Assistant Laboratory for Experimental Psychology Faculty of Philosophy University of Belgrade, Serbia and Montenegro
Fall 1999	Statistical Techniques for undergrads majoring in Psychology
Spring 2000	Introductory Psychology for undergrads majoring in Psychology
Fall 2001	Grader, Spring Mathematical psychology
Fall 2003	Grader, Fall Mathematical psychology
Fall 2002	Advanced Statistical Analysis I
Spring 2003	Advanced Statistical Analysis II
Fall 2004	Advanced Statistical Analysis I
Fall 2005	Advanced Statistical Analysis II

Professional Societies

Society for Mathematical Psychology

Research Assistantship

1994-1999	Laboratory for Experimental Psychology Faculty of Philosophy

University of Belgrade, Serbia and Montenegro

Reviewer

Ad hoc reviewer for the
Perception and Psychophysics,
Journal of Mathematical Psychology
Journal of Experimental Psychology: Human Perception and Performance
Psychology Review

Collaborations:

- The application of mental processing decomposition testis's in autism in collaboration with, Shannon Johnson and Dr. Julie Stout, Indiana University. *Clinical & Cognitive Neuroscience Research, of IU Bloomington, Psychology.*

- The application of intervention analysis in time series, on single subject ANOVA test. Dr. Jerome Busemeyer, Indiana University:

- Identification of cognitive structure in dyslexic children (1998) with Rosana Brakus, *Institute for experimental phonetics and speech pathology*, Belgrade, Serbia and Montenegro

Publications

Fific, M., (2005) Emerging holistic properties at face value: Assessing characteristics of face perception. *Unpublished Ph.D. thesis*.

Townsend, J.T., Fific, M., & Neufeld, R.W.J. (in press). Assessment of mental architecture in clinical/cognitive research. In T.A. Treat, R.R. Bootzin, T.B. Baker (Eds.), Psychological clinical science: Papers in Honor of Richard M. McFall. Mahwah, NJ: Lawrence Erlbaum Associates.

Fific, M., Townsend J. T. & Eidels A. (2005) Simple Visual Search is not Always Parallel (or Serial either): Evidence from Systems Factorial Technology (submitted):

Townsend, J. T., Fific M. (2004) Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics*. Vol 66(6), pp. 953-962. [PDF]

Townsend, J. T., Fific M., Assadi, A. (2003). *General recognition theory and probabilistic perceptual separability on simple cognitive surfaces.* Paper presented at the Fechner Day 2003, Larnaca, Cyprus.

Townsend, J. T., Fific M. (2001). *Representation and process in defining holisms partism: contributions from general recognition theory and stochastic cognitive process theory.* Paper presented at the Fechner Day 2001, Leipzig, Germany.

Fific M., (1999).Temporal factors in short-term memory search, part I: An introduction. *LEP report, 72.*
Fific M., (1999).Temporal factors in short-term memory search, part II: Organization of memory and memory search. *LEP report, 73.*
Fific M., (1999).Temporal factors in short-term memory search, part III: Recognition and reproduction. *LEP report, 74.*
Fific M., (1998). Selective attention and information processing in short-term memory. *Paper presented at the symposium of Ergonomics, Belgrade 1998, 19-23.*
Duzdevic N. & Fific M. and Brakus R., (1998). Education of mentally preserved children with cerebral palsy based on the level of their achievements. In monograph: Cerebral palsy, Belgrade 1998, 209-212.
Fific M., (1998). Dynamics of Serial Position Change in Probe-Recognition Task. *LEP report, 64.*
Fific M., (1997). Determination of stimulus-pair order in short-term memory. *LEP report, 34.*
Fific M., (1996). Modification of binary response and short-term memory processing. *Psihologija, 29, 2-3, 311-331.*
Fific M., (1996). Modification of pre-probe delay in the short-term memory recognition task. *Psihologija, 29, 2-3, 331-353.*

Recent Presentations:

2005. November 10, Object, Perception, Attention, and Memory (OPAM), Toronto Ontario: *Emerging holistic properties at face value: Assessing characteristics of face perception*. Fific & Townsend [Poster]
2005. September 21, Cognitive Lunch, Indiana University: *Emerging holistic properties at face value: Assessing characteristics of face perception.*
2005. August. 5, Mathematical Psychology, Memphis University: A processing model of holistic/configural face perception. Fific & Townsend
2004. November 19, Minneapolis 45th Annual Meeting of the PSYCHONOMIC SOCIETY: *Perceptual and Cognitive Processes in Autism Spectrum disorder: New Perspectives on Global-Local Processing*. Johnson, Blaha, Fific & Townsend [Poster]
2004. November 19, Minneapolis 45th Annual Meeting of the PSYCHONOMIC SOCIETY: *Proposed Model for Efficiency Variability in a Visual Search Task.* Fific & Townsend [Poster]
2004. December 17, Cognitive Lunch, Indiana University: *Proposed Model for Efficiency Variability in a Visual Search Task.* [PP presentation]
2004. July 30, Society for Mathematical Psychology, University of Michigan, Michigan: *Proposed model for efficiency variability in a visual search task*. [PP presentation]
2004. June 12, HML, Ohio: *Proposed model for efficiency variability in a visual search task* [PP presentation]
2003 July 26 Mathematical Psychology, Utah University: *Properties of visual search task on two items revealed by the Systems factorial methodology*

2003 June Laboratory for experimental psychology, Belgrade University, Invited talk,: *A test for holistic and analytic face perception*

2002. July 26. Mathematical Psychology, Ohio University: *Multiple-channel processing in short-term memory*

2002. June 10, Laboratory for experimental psychology, Belgrade University, Invited talk: *Revealing properties of mental organization using system factorial technology in memory and visual search*

2002. April 7, HML, Ohio,: *Do different processing strategies (serial vs. parallel) operate on first and last memorized items?*

2002. February 7, AIC Boulder: *Parallel vs Serial Processing and Individual Differences in Visual & Memory Search Tasks*

2003. March 28 HML, Purdue University,: *Properties of visual search task on two items revealed by the Systems factorial methodology*

2001. November 28, Cognitive Lunch., Indiana University: *Multiple-channel processing in short-term memory*

2001. November 17, Psychonomics Orlando: *Strong Evidence on Parallel vs. Serial Processing in the Sternberg Paradigm. (Presented by Jim Townsend)*

2001. July 26 Mathematical Psychology, Brown University: *Revealing mental architecture in memory search task by systems factorial technology*

2001. June 1, Laboratory for experimental psychology, Belgrade University, Invited talk: *Revealing properties of mental organization using system factorial technology*

2001. March 3, Hoosier Mental Life: *Revealing mental architecture in memory search task by systems factorial technology*

2001. March 28, Cognitive Lunch, Indiana University: *Revealing mental architecture in memory search task by systems factorial technology*