AGE OF ACQUISITION
AND FAMILIAR FACE RECOGNITION

Rachel Richards

Submitted for the degree of PhD

The University of York
The Department of Psychology

May 2006

ABSTRACT

The processing of information learned early in life is superior compared with that learned later. Two aspects of this age of acquisition (AoA) effect in face recognition were investigated. The first sought to explore the relationship between AoA and distinctiveness in order to offer a refinement of the mechanisms underpinning AoA and shed light on associated loci. The second recruited AoA as a tool with which to assess the relationship between identity and gender processing and constrain mechanisms purported to subserve this relationship. The mutual investigation of AoA and face processing sought to derive implications for each area of research.

The Pilot Experiment and Experiments 1 to 3 led to the conclusion that AoA and distinctiveness have an additive relationship. Although the results of these familiarity decision tasks were not entirely consistent, inclinations towards additivity were considered more persuasive than various interactions recovered under suboptimal circumstances. The independence of the variables indicated that they are subserved by different mechanisms which require no integration to derive theories of AoA. Moreover, distinctiveness and AoA were tentatively ascribed to serially-arranged processing stages, interpreted with reference to contemporary models of face recognition.

On the basis of Experiments 4 to 10 it was concluded that mechanisms underpinning identity and gender processing are not entirely independent. An effect of distinctiveness on gender decisions, independent of gender typicality effects, indicated that male and female faces systematically colonise multidimensional face-space (MDS). As such, MDS may provide a locus for interfacing identity and gender processing. Effects of AoA on gender decisions emerged under conditions where it is hard to glean gender from a visual analysis of gender cues alone. Coupled with an effect of gender typicality on familiarity decisions it was concluded that the relationship between identity and gender processing may be one of mutual influence, best described by parallel-route mechanisms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Thank you to Andy Ellis for giving me the opportunity to carry out this research and offering me invaluable advice, encouragement and wise words. His skills at keeping panic at bay are second to none. Also, a huge thanks to Andy Young and Philip Quinlan for their insights and suggestions. For keeping me sane over the years, I thank my friends, Emily, Leanne, Liz, Marilena and Michelle, and my family. For his continued support and dedication, I will always be grateful to Shaun.

Thank you also to Julia Rayner for providing me with the stimuli to test out my ideas in the early days, and to PA Photos for allowing me access to images which contributed to my database.

DECLARATION

This thesis contains original work completed by Rachel Richards under the supervision of Professor Andrew Ellis.

The data reported in Experiments 2 and 3 have been presented at the BPS Cognitive Section Conference 2004 in Leeds:

Richards, R. & Ellis, A.W. (2004). Is there an interaction between age of acquisition and distinctiveness in familiar face recognition? Poster presented at the BPS 2004 Cognitive Section Conference, Leeds.

CHAPTER 1 – INTRODUCTION AND REVIEW OF THE LITERATURE

1.1 Introduction

This thesis is concerned with an exploration of cognitive operations associated with age of acquisition (AoA) in the recognition of familiar faces. The AoA effect refers to the superior processing of information learned early in life compared with that learned later. Lewis (1999a) stated that whilst there may be an abundance of evidence concerning the existence and size of AoA "mechanisms by which effects of AoA are suggested to take place are few and lack a framework than can be implemented in order to derive predictions" (p.25). Two aspects of the AoA effect in face recognition are pursued in this thesis. The first seeks to ascertain the nature of the relationship between AoA and facial distinctiveness and concomitantly offer a refinement of the mechanisms underpinning AoA (Section 1.1.1). The second employs AoA as a tool to with which to assess the relationship between identity and gender processing and constrain proposed explanations (Section 1.1.2).

1.1.1 Exploration of the Relationship Between AoA and Distinctiveness

The distinctiveness effect refers to the superior processing of distinctive or unusual faces, compared with typical or more average faces. However, research has shown that effects are attenuated in childhood emerging only as a function of experience with faces. It is feasible that AoA effects are modified by variables, such as distinctiveness, that are dynamic in their effect over time: the childhood attenuation of distinctiveness might be duly imprinted in the representation of faces acquired early, ensuring that effects of distinctiveness are restricted to the processing of faces acquired later.

An interaction between AoA and distinctiveness would serve to elaborate existing theories of AoA and distinctiveness: this thesis couches explanations in terms of A.

Ellis and Lambon Ralph's (2000) neural plasticity account of AoA and Valentine's (1991) multidimensional face-space (MDS).

Reviews of the seminal face recognition models are considered in Section 1.2.1. Section 1.2.2 discusses models of face representation, more specifically, Valentine's (1991) MDS. These facilitate theoretical interpretations of the AoA and distinctiveness literature in addition to the ensuing results. A selection of the distinctiveness literature is reviewed in Section 1.3. The mnestic basis of the variable is initially established, essential if it is to interact with AoA. A review of the temporal dimensions of distinctiveness includes developmental accounts and the relationship between distinctiveness and facial ageing: a positive relationship between distinctiveness and perceived age sparked an investigation of the interaction between AoA and distinctiveness with dated images in addition to current images. Section 1.4 presents AoA literature, tracking its conception as a stage-dependent process to a general by-product of the way in which information is stored and accessed in the brain. Particular emphasis is placed on contemporary theories, including A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis.

1.1.2 Exploration of the Relationship Between AoA and Gender Processing

AoA is recruited as an investigative tool, facilitating exploration of the relationship between identity and gender processing. Identity and gender were originally believed to be processed independently, reflected by their functional separation in Bruce and Young's (1986) model of face recognition. However, evidence that the cognitive operations are not entirely independent has accumulated and various degrees of integrality of processing proposed. Conditions under which AoA effects emerge during gender processing tasks are manipulated to constrain interpretations of this relationship.

A by-product of this exploration is the assessment of AoA effects in tasks which measure the variable indirectly. Gender processing tasks allow AoA effects to be monitored without tapping explicit recognition processes, and thus, performance is less subject to error.

Literature investigating the relationship between identity and gender processing is reviewed in Section 1.5. The various conceptualisations of the relationship and theoretical interpretations open up options for the interpretation of findings. The feasibility of Valentine's (1991) MDS as an interface between identity and gender processing stems from a consideration of studies which provide evidence that gender is coded in MDS. Sections 1.2.1 (models of face recognition) and 1.2.2 (models of face representation) contextualise Section 1.5 and enhance theoretical interpretations of the literature and present findings.

1.2 Models of Face Recognition and Representation

Research into face recognition, tackled from a cognitive stance, seeks to clarify the different operations involved in face processing and how they interact. Models of face recognition, and the associated process of representation, are able to accommodate a wide range of psychological phenomena which inform an increasingly detailed understanding of mechanisms underpinning recognition and representation. Models of recognition and representation are reviewed to contextualise subsequent reviews of AoA, distinctiveness, and the relationship between identity and gender processing.

1.2.1 Models of Face Recognition

Recognition is defined as the product of a match between a perceived stimulus and a stored representation held in memory. Face recognition research has pursued disparate routes of investigation with the cognitive processes following perception of the stimulus modelled largely in isolation from the perceptual processing itself. This section documents cognitive (Section 1.2.1.1) and perceptual (Section 1.2.1.2) models of familiar face recognition, culminating in a review of a more comprehensive model (Section 1.2.1.3).

In addition to the behavioural studies which motivated their creation, evidence supporting the architecture of face recognition models has converged from neuropsychological research and brain imaging studies. Research with implications for the hierarchic structure of the recognition route and the heterarchic nature of identity and gender processing routes is reviewed in Section 1.2.1.4.

1.2.1.1 Cognitive Models

Until the late 1970s, literature on face processing was eclectic and lacking the coherence necessary for deriving theoretical models. With the advent of cognitive approaches, theoretical models have since organized and fuelled research. Developed from an earlier model presented by Hay and Young (1982), the Bruce and Young (1986) model of face recognition (Figure 1.1) is perhaps the most well-known product of this spate. Inspired by evidence from behavioural experiments, studies of everyday recognition errors and neuropsychology, the model accounts for the sequentially dependent cognitive operations involved in identification, housed by the model's hierarchic architecture. In addition, heterarchic architecture accommodates the functionally distinct processes of identification, expression analysis, facial speech analysis and directed visual processing; the latter refers to the selective and strategic encoding of facial information. Whilst the model is largely agnostic with respect to perceptual processing, a structural encoding stage specifies the products of perceptual processing: viewer-centred descriptions provide information for the analysis of facial expressions, facial speech and directed visual processing; expression-independent descriptions also feed directed visual processing in addition to providing information amenable for storage in face recognition units (FRUs) which kick start the identification route.

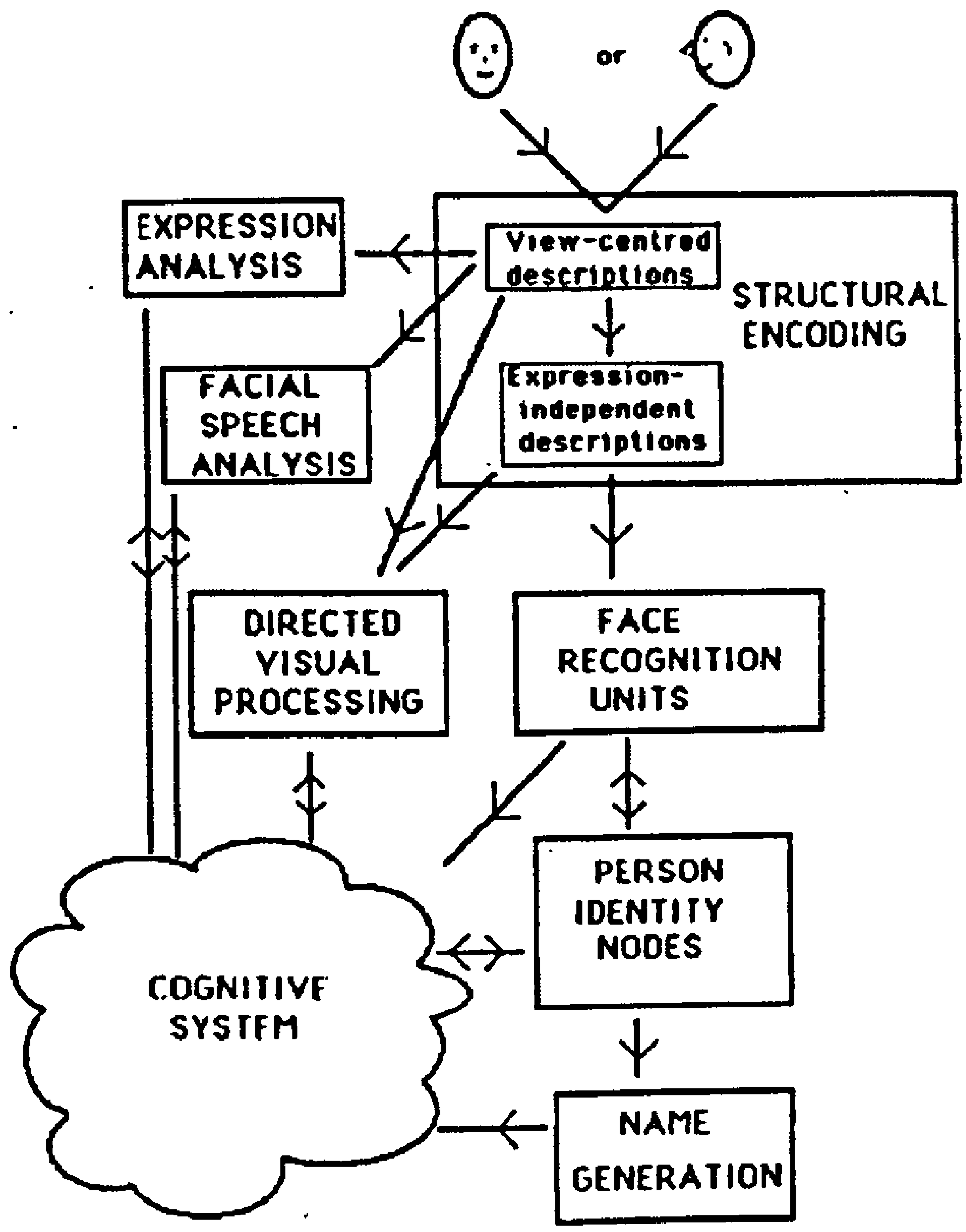


Figure 1.1. *Bruce and Young's (1986) Functional Model of Face Recognition.*

FRUs contain abstract structural codes, one of which exists to represent each known face. Recognition occurs when there is a match between the structural encoding unit input and the description stored in FRUs. Person identity nodes (PINs) are accessed via FRUs and contain identity-specific semantic information. They are modality-independent and thus mark "the point at which person recognition as opposed to face recognition is achieved" (Bruce & Young, 1986, p. 312). PINs provide a gateway to the cognitive system, although the precise nature of this relationship is unclear. Whilst Bruce and Young do not explicitly separate semantic information from the PINs, it is not clear whether semantic information is stored in the PINs themselves, or accessed via the PINs from associative memory held in the cognitive system. Name retrieval is contingent upon identity-specific semantic information.

The cognitive system receives input from each of the heterarchically arranged processing routes and in turn, it can influence each of these functional components. Using information from structural encoding, expression analysis, directed visual processing and FRUs, the cognitive system generates visually-derived semantic codes. These codes contain information that can be gleaned from familiar and unfamiliar faces alike, for example, gender, age and personality attributions. However "future studies may allow the separation of visually derived semantic codes into distinct types, produced by different routes" (Bruce & Young, 1986, p, 313). The authors remark that gender and age judgments may require different processes from those involved in judging honesty for example. It is interesting to note the nebulous basis of gender processing.

To provide a working implementation of Bruce and Young's (1986) model of familiar face recognition, Burton, Bruce and Johnston (1990) employed interactive activation competition (IAC) architecture to 'extend the microstructure' of the hierarchic identification route. This localist connectionist model (Figure 1.2) comprises active units connected by modifiable links. Units receive activation which is either provided externally by the experimenter or internally from other units in the net. They are organized into pools and connected within pools by bidirectional inhibitory links. Associated units between pools are connected with bidirectional excitatory links. The main thrust of Bruce and Young's identification route is preserved by three pools of units corresponding to FRUs, PINs and semantic information units (SIUs). Note that Burton et al. (1990) explicitly separate semantic information from the PINs, defining PINs as nodes which allow access to SIUs. Another theoretical modification entailed relocating familiarity decisions from the FRUs to the PINs; a modality-independent locus. Hence, removing the need for modality-dependent loci, each responsible for separate face, name and voice familiarity decisions.

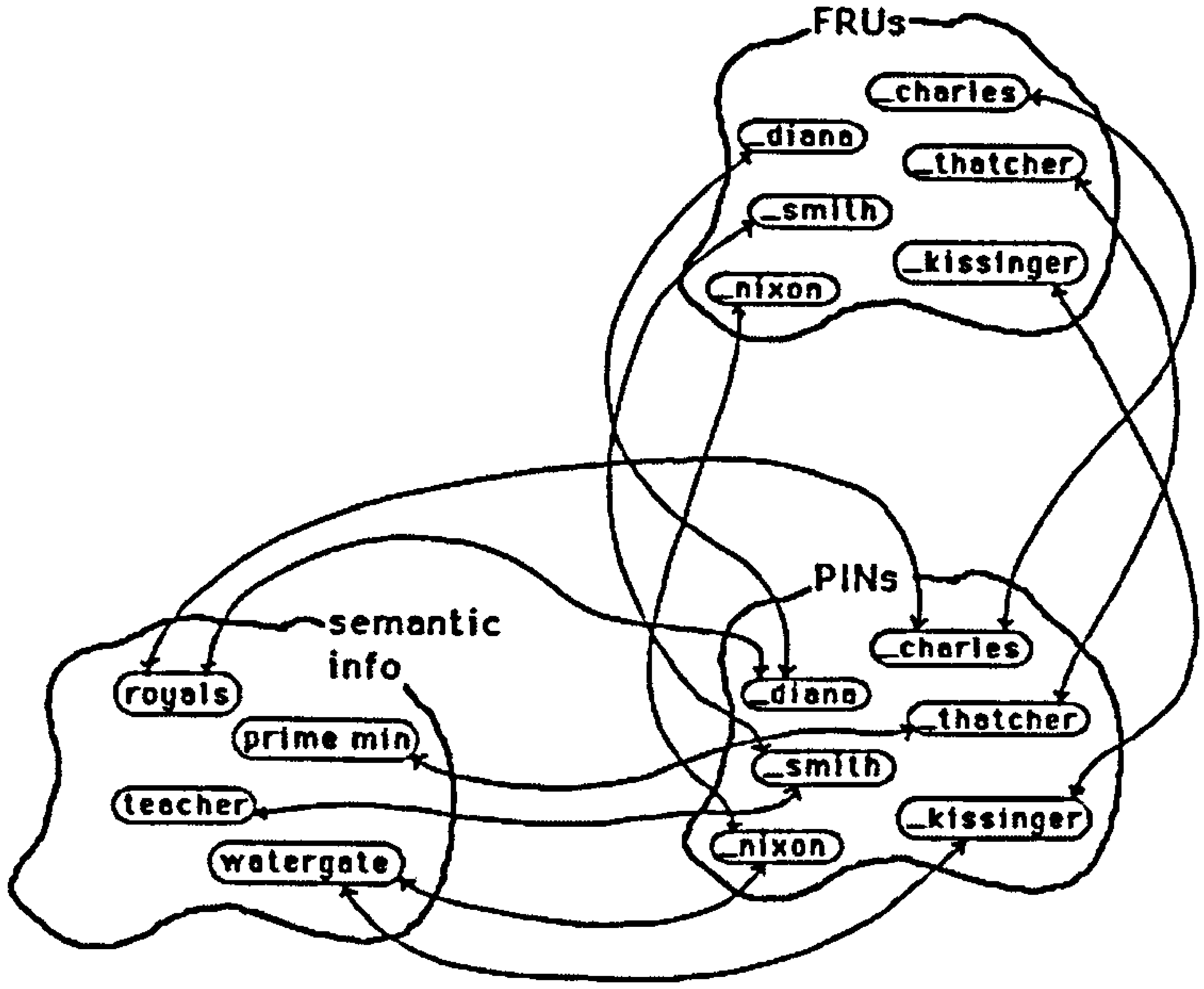


Figure 1.2. *Burton, Bruce and Johnston's (1990) IAC model of Familiar Face Recognition.*

The model successfully simulated empirical phenomena including semantic and repetition priming. With the addition of a simple front-end to the model, loosely analogous to feature-based structural encoding units, distinctiveness effects were also simulated. However, the model does have its theoretical limitations. In addition to elucidating cognitive operations involved in face processing, the cognitive approach seeks to understand how these operations interact with one another. The authors conceded that to explore the question of additivity and interactions, the size of the main effects need to measured. However, the model is not able to make claims regarding the mapping between its dependent variable (the number of activation cycles required in order to reach an arbitrary threshold level of activation) and reaction times.

1.2.1.2 A Perceptual Model

Hancock, Burton and Bruce (1996) examined the feasibility of principal components analysis (PCA) of face images (Kirby & Sirovick, 1990; Turk & Pentland, 1991) as a psychologically plausible front-end to face recognition models. Unlike previous image coding schemes which nominated various primitives as candidates, PCA codes image intensity information acknowledged for its role face recognition (e.g. Bruce et al., 1993a). This technique also compresses data by reducing the number of dimensions which code variability between faces. For any given set of faces, a one-dimensional array of pixel values is derived per face. Correlations between these images are calculated and eigenfaces are extracted; eigenfaces code deviation from the mean. Eigenfaces extracted early in the process code maximum variance between the images, thus include information carrying gender. Later eigenfaces carry finer detail, such as identity.

Hancock et al. (1996) reasoned that the use of PCA to derive a set of dimensions along which faces vary, affords a neat implementation of psychological theories such as distinctiveness, which are subserved by similar mechanisms. This implementation was put to the test. Participants were asked to rate a set of images for distinctiveness followed by a recognition memory task. Whilst distinctiveness was correlated with false positives (misidentified faces) and hits (correctly identified faces), these two performance measures were not correlated. The images were subsequently subjected to PCA to establish whether or not this technique could explain these orthogonal bases of distinctiveness. Through a variety of measures, the authors found that early eigenfaces predicted false positives and late eigenfaces predicted hit rate. Given that early components code general information and later components code individual variation, the authors speculated that hits are largely idiosyncratic in origin and false positives, a product of the similarity of a face to the general population.

Its ability to isolate independent statistical properties corresponding to hits and false positives provides compelling evidence for the psychological plausibility of PCA as a perceptual extension to cognitive models. However, the authors close with the caveat that the present results may be a general consequence of any image-based statistical technique.

1.2.1.3 A More Comprehensive Model

A model of familiar face recognition with integrated perceptual and cognitive components was devised by Burton, Bruce and Hancock (1999), see Figure 1.3. The perceptual front-end recruited PCA of face images and the cognitive back-end was implemented as IAC architecture akin to that developed by Burton et al. (1990). The model sought to promote understanding of psychological phenomena outside the range of models which focused solely on cognitive or perceptual components. To avoid repetition of sections 1.2.1.1 and 1.2.1.2, the description of the architecture is restricted to the 'join' between the two components.

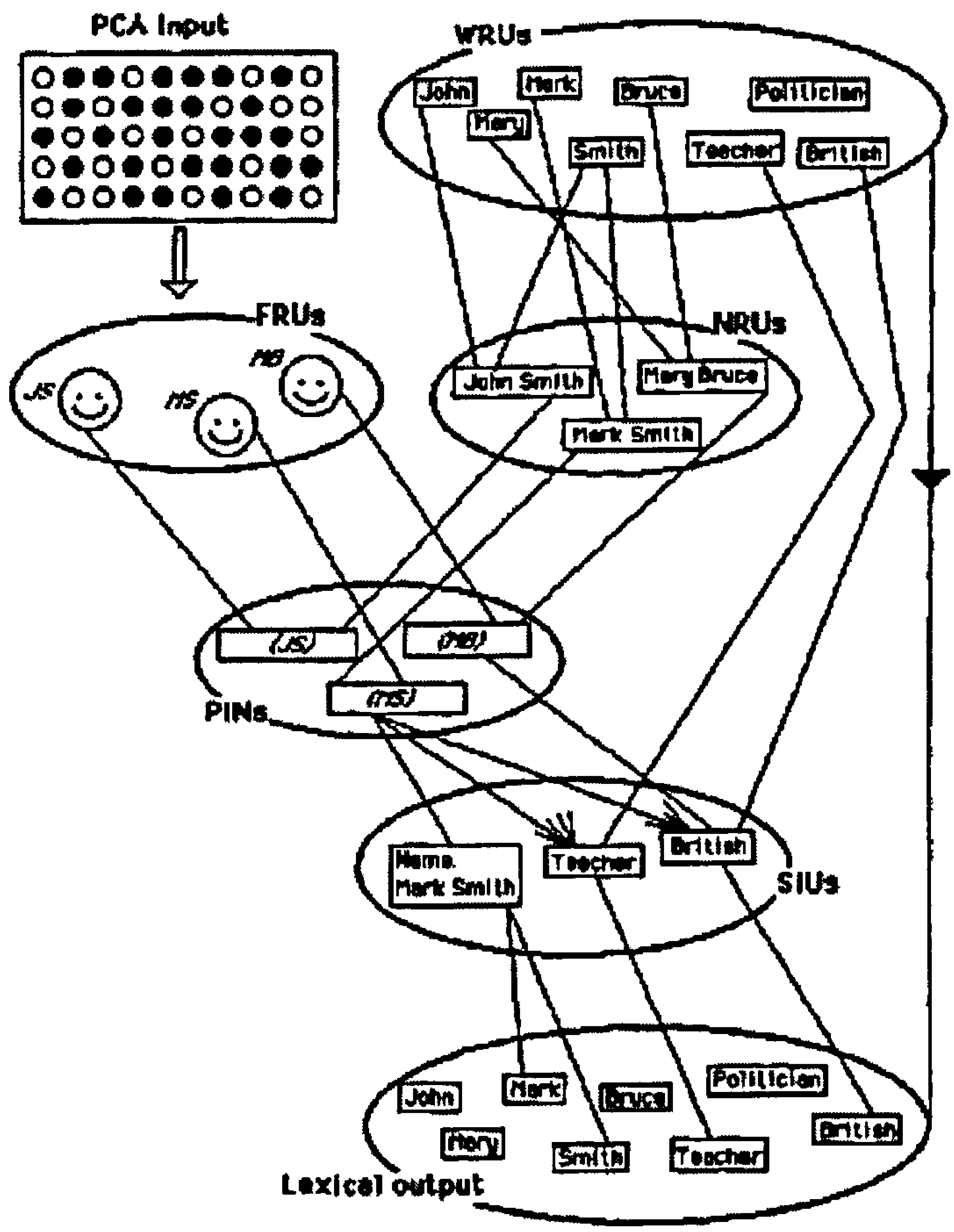


Figure 1.3. *Burton, Bruce and Hancock's (1999) Model of Familiar Face Recognition.*

The model was constructed to recognise 50 people, thus eigenfaces were derived from 50 male full-face neutral expression photographs. The perceptual front-end was tied to the cognitive component through FRUs, via PCA input units which are connected to FRUs according to the signature of each face. A signature refers to a set of values corresponding to the eigenface dimensions for that face. The authors describe the connection: "If a face has positive value on component 1, and negative on component 2, a link of strength +1 is constructed between PCA unit 1 and that FRU, and a strength of -1 is constructed between PCA unit 2 and that FRU. This procedure is followed for all PCA units and all FRUs" (p. 14). To allow for recognition in additional domains, the IAC architecture extends Burton et al.'s (1990) implementation through the addition of name recognition units, word recognition units and lexical output units.

The model simulated psychological phenomena arising from an interaction between perceptual and cognitive aspects, such as cross-modal priming and distinctiveness effects. In addition, effects emulated by previous models which had unspecified perceptual inputs, such as cross-domain semantic priming and image-dependent repetition priming effects, persisted with a PCA front-end. The model also extended the predictive power of PCA. Research carried out by Hancock et al. (1996) had been effectively restricted to investigating the psychological plausibility of PCA as a front-end to unfamiliar face processing.

1.2.1.4 Models of Face Recognition: Neuropsychological Support for Hierarchic and Heterarchic Structures

Neuropsychological research has pinpointed selective deficits associated with each recognition stage, strengthening the notion that each is sub-served by different cognitive processes. To extend these findings, recent imaging studies sought to clarify whether familiarity is signalled at FRUs, a modality-dependent locus (Bruce & Young, 1986) or at PINs, a modality-independent locus (Burton et al., 1999; Burton et al., 1990).

Whereas face processing is associated with activity in the lateral fusiform gyrus, recognition of known faces is associated with additional activity in the anterior temporal region. Using functional magnetic resonance imaging (fMRI), Leveroni et al. (2000) found that famous faces were associated with anterior temporal gyrus activity compared with newly familiarized and unfamiliar faces in an 'old' (seen before) versus 'new' task. The authors inferred that these regions must be associated with the representation of semantic knowledge.

The anterior temporal region could either be conceptually synonymous with the PINs (Bruce & Young, 1986) or SIUs (Burton et al., 1990) depending on the functional interpretation of these stages. Shah et al. (2001) used the premise, 'irrespective of function PINs are modality independent', to explore the neural substrates of familiarity decisions and identity-specific semantic information. During a passive observation task interleaved with subsidiary tasks designed to maintain attention, activity elicited by familiar and unfamiliar faces and voices was measured by fMRI. Regardless of modality, the retrosplenial cortex was activated by familiar items. Shah et al. concluded that because the region associated with semantic knowledge (anterior temporal region) showed no activity, retrosplenial activity corresponded to familiarity decisions. The modality-independent locus of familiarity decisions renders Burton et al.'s (1990) definition of PINs the most appropriate description of the data.

As far as the heterarchic structure of Bruce and Young's (1986) model is concerned, whereas neuropsychological double dissociations largely support the functionally distinct processing routes, the separability of identity and gender processing is sustained by a single-dissociation (Tranel, Damasio & Damasio, 1988). Whilst Bruce, H. Ellis, Gibling and Young (1987a) point out that patients may judge gender on the basis of superficial characteristics, thus masking an inability to judge gender, the afore-mentioned double dissociations are largely underpinned by imaging studies (e.g. Calvert et al., 1997), but mixed research underpins the single dissociation.

In a positron emission tomography (PET) study, Sergent, Otha and Macdonald (1992) found that different anatomical regions were activated during gender classification and familiarity decision tasks. However face familiarity confounded the tasks. In a subsequent PET study, where the confound was removed (Dubois et al., 1999), identical brain activity, predominantly in the fusiform face area, was triggered in response to gender and familiarity judgments. This spurred the conclusion that this region could provide a neat correlate to the structural encoding stage of cognitive models. However, in a recent review Haxby, Hoffman and Gobbini (2000) proposed that the fusiform gyrus mediated the processing of invariant information. Hence, unique identification is processed in a package with other invariant information, such as gender

Dubois et al. (1999) also uncovered a processing difference between newly familiarised and unfamiliar faces in their gender classification task. This finding was replicated by Rossion, Schiltz, Robaye, Pirenne and Crommelinck (2001) who carried out a gender classification task with morphed familiarised and unfamiliarised faces: the resultant processing differences related to the dissociation between faces perceived as familiar and those perceived as unfamiliar. Both studies reported a decrease in brain activity associated with familiarised faces compared with unfamiliar faces. As a result, Dubois et al. and Rossion et al. (2001) deemed a strict independence of identity and gender processing untenable.

1.2.2 Models of Face Representation

Whilst there has been a degree of convergence in models of face recognition, theories concerning the representations derived and stored to mediate this process are less cohesive. Nevertheless, Valentine's (1991) multidimensional face-space (MDS) has emerged as a pivotal contender. Section 1.2.2.1 offers a review of the structure of MDS, followed by an elaboration of this framework as the dichotomous norm-based and exemplar-based interpretations in Section 1.2.2.2.

Two major issues associated with the operation of the model are also subject to scrutiny. The accommodation of invariance is discussed in Section 1.2.2.3, whereas Section 1.2.2.4 reviews the metric used to evaluate similarity between percept and stored representation.

1.2.2.1 Basic Structure of MDS

Valentine (1991) proposed that a location in Euclidean space provides a metaphor for the representation of a face. MDS (Figure 1.4) comprises sufficient dimensions to represent any physiognomic feature used to discriminate faces, but the nature of these remains unspecified. The origin of the space represents the average value of the population on each dimension, thus it is assumed that representations form a multivariate normal distribution. Consequently, the origin is assumed be the point of highest density and density decreases as a monotonic function of the distance from the origin. The differential density manifests itself as effects of distinctiveness: whereas typical faces are clustered in central densely populated areas, distinctive faces colonise the sparsely populated periphery. Recognition of distinctive faces, relative to their typical counterparts, is facilitated owing to the reduced local density. Fewer competing exemplars negate competition.

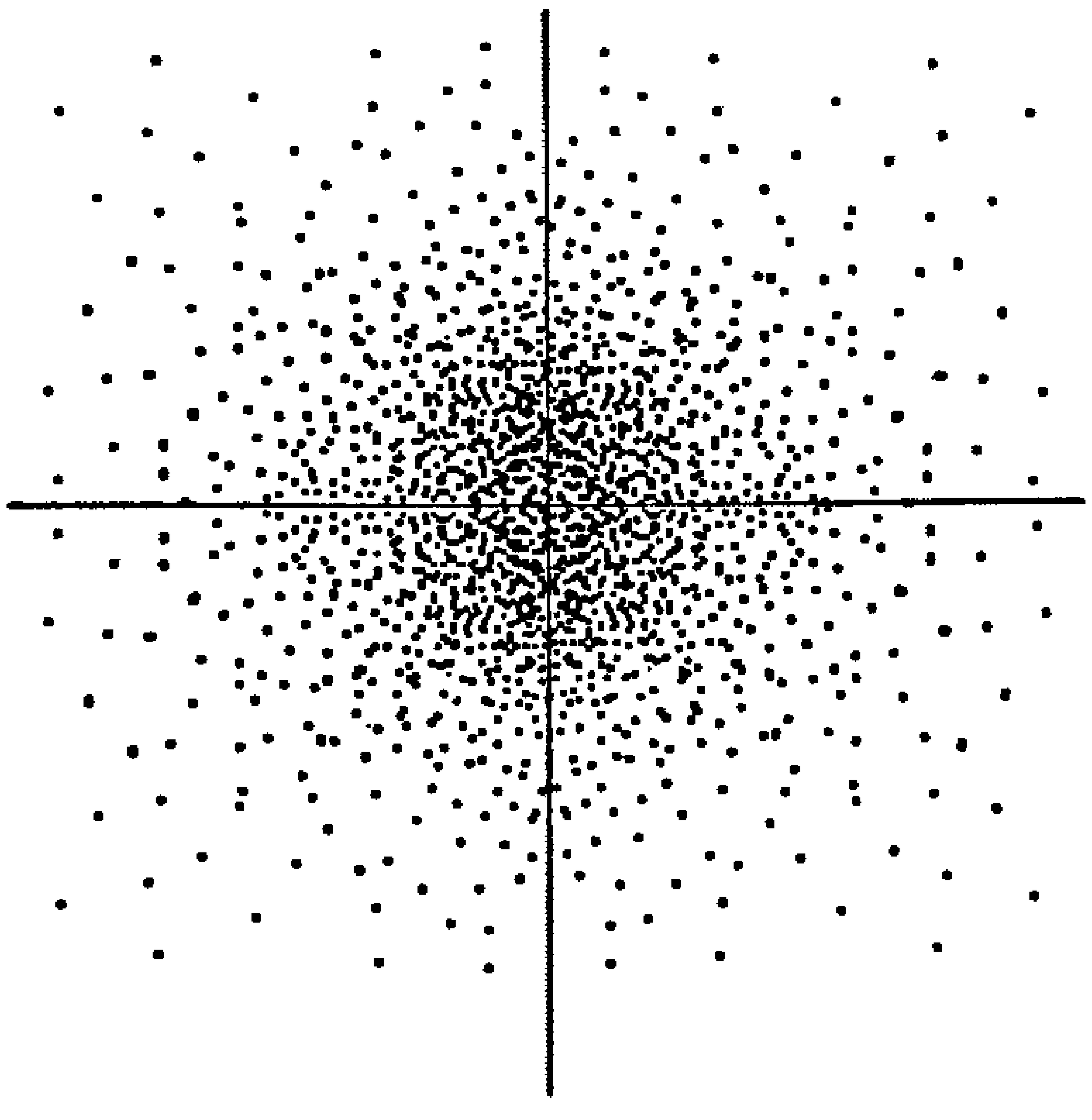


Figure 1.4. *A two-dimensional representation of Valentine's (1991) multidimensional face-space.*

Bruce, Burton and Dench (1994) investigated the physical correlates of distinctiveness to elucidate the nature of the dimensions structuring MDS and provide evidence that faces are normally distributed. Bruce et al. (1994) collected anthropometric measures from 175 photographs of unfamiliar full-faces and complex measures, including 3D protuberance and ratios, measured from these full-faces and their profiles. The sum of the modulus of the z-scores for each measurement indexed physical deviation from the mean. Multiple linear regression analysis regressed anthropometric and the complex measures, one at a time, and together, on to distinctiveness ratings. Regressions which included eccentricity measures derived from both sets of measurements accounted for more of the variance than eccentricity derived from one set alone. Given that physical measurements predicted distinctiveness at all, led the authors to conclude that the encoding of faces along explicit dimensions is conceivable. Nevertheless, Bruce et al. concede that facial attributes more complex than those measured may account for the representation of faces. Indeed, they suggested that PCA may offer a worthwhile interpretation of MDS dimensions. Significant correlations between distinctiveness ratings and eccentricity led Bruce et al. to conclude that distinctiveness reflects eccentricity from a mean. Consequently, the distribution of faces in MDS was assumed to be consistent with a normal distribution or a centrally-clustered distribution.

Despite the acknowledgement that the majority of faces are not typical (Vokey & Read, 1995), MDS is characterised by its multivariate normal distribution. In an attempt to resolve this paradox, Burton and Vokey (1998) explored the definition of typicality and discrepancies which arise when MDS is depicted as a bivariate space, rather than the multivariate space intended.

Typicality can be defined with regard to local density (the proximity between a face and its immediate neighbours), which is highest at the centre of space. Furthermore, it is commonly assumed that the majority of faces occupy the region of highest density. However, as the majority of faces are not considered very typical, it appears that the majority of points, rather counterintuitively, cannot fall in this high density region of space. This notion makes more sense when MDS is considered as a multivariate space, rather than the bivariate space in which it is commonly depicted. Burton and Vokey demonstrated that when MDS is conceptualised as a bivariate space, with values of each dimension varying normally around the central tendency, the frequency of faces located at any given point appears to be a monotonically decreasing function of squared distance from the centre. However, this is not invariant as the number of dimensions increases. The authors demonstrated that when dimensionality is greater than 2, the majority of faces are no longer found at the centre of space. Moreover, the mode moves progressively further away from the centre as dimensionality increases. This observation can explain why local density is highest at the centre of a space with normally distributed dimensions, and yet the majority of faces are not very typical. For the sake of parsimony, a bivariate space is recruited to aid theoretical interpretation in this thesis. However, it is acknowledged that face space is multivariate and as such properties of MDS vary accordingly.

Wickham, Morris & Fritz (2000) analysed the distribution of faces rated for distinctiveness as a function of rating instructions. 'Traditional' rating scales asking participants to decide how distinctive a face is, and 'deviation' scales asking participants to rate faces according to how different they are from an average face, were administered. A normal distribution of ratings along the typical-distinctive dimension was produced by the 'traditional' scale supporting Burton and Vokey's (1998) computation. However, the 'deviation' scale yielded a positive skew of ratings towards the typical end of the dimension. Wickham et al. (2000) postulated that the style of instructions is responsible for discrepancies between conventional accounts and Burton and Vokey's contention. In order to make the task easier, participants had to make forced-choice distinctiveness judgments to pairs of faces. A distribution akin to that proposed by Burton and Vokey emerged. As a reconciliation, Wickham et al. suggested that participants may rate distinctiveness for reasons other than the combined influence of all dimensions. However, the universal reliability of distinctiveness ratings reported across studies suggests that idiosyncratic use of dimensions is an unlikely source of judgments.

1.2.2.2 Norm-Based Versus Exemplar-Based Interpretations

Valentine (1991) distinguished norm-based and exemplar-based interpretations of MDS. The norm-based conceptualization dictates that faces are encoded as vectors which reference their deviation from a prototype. Similarity between two exemplars is dependent on the distance separating them and the length of the vector linking exemplar and prototype. Conversely, exemplar-based accounts provide no role for an abstracted norm and consider faces to be represented as points. In this case, similarity is determined solely by inter-point distance. Whereas exemplar-based models explain distinctiveness effects seen in other-race faces more parsimoniously than norm-based models (Valentine & Endo, 1992), norm-based models are traditionally recruited to account for the recognition advantage imparted by caricatured faces over their veridical counterparts (Rhodes, Brennan & Carey, 1987).

In order to derive a unified model, Lewis and Johnston (1999) reasoned that "it is either necessary to show how a norm-based model can account for race effects, or to demonstrate how an exemplar-based model can account for the caricature advantage" (p.10). In pursuit of the latter, they devised their Voronoi model, founded on the premise that encoded representations are bounded by identity regions. The perimeter of these regions bisects inter-point distance. The authors demonstrated that the caricature advantage is an emergent property of representing faces by regions because the centre of the identity region (the point of optimum activation) will be further from the centre of MDS, hence in a lower density area, than the face which formed the region (the veridical image). Hence, caricatures optimally activate the region, compared with the veridical representation. Unless otherwise stated, forthcoming references to MDS assume an exemplar-based conceptualisation, owing to its ability to account for a wide range of phenomena.

1.2.2.3 Nature of Representations

Faces can be identified despite changes in lighting, view, expression and transformations due to age, therefore models must take into account demands imposed by the requirement of invariance. Newell, Chiroro and Valentine (1999) investigated view-based and individual-based interpretations of MDS in order to explain view-point invariance. The view-based account hypothesises that different views inhabit different subspaces and the location of a representation remains invariant across subspaces. Invariance is imposed by tagging together view-specific representations of the same individual across the subspaces. This theory is mechanistically analogous to Craw's (1995) manifold model of object and face recognition which posits that different views of a face are integrated across subspaces in an 'identity manifold'. Newell et al.'s individual-based account hypothesised that different views of the same individual cluster in MDS, correlated by close temporal proximity. This purely instance-based interpretation considers representations to be points rather than manifolds, analogous to FRUs (Bruce & Young, 1986) which "respond when any view of the appropriate person's face is seen" (p. 311-312).

In a series of experiments, Newell et al. (1999) reported that distinctiveness and view did not interact. This afforded support for the view-based approach which supposed that because generalisation to different views occurs across sub-spaces, recognition should be affected in a similar way regardless of distinctiveness. The individual-based account was not supported because it predicted that generalisation across views is affected by distinctiveness, with an increase in encoding error affecting the densely populated typical faces most adversely. Although Lewis and Johnston's (1999) Voronoi model is an individual-based manifestation of MDS, Valentine (2001) reconciled the Voronoi model and manifold accounts. He hypothesised that dimensions coding identity partition the space into identity regions, but dimensions coding transformations such as view, lighting, expression and age may be captured across manifolds.

1.2.2.4 Nature of Similarity Metric

Recognition requires an evaluation of the similarity between a perceived face and stored representations, but "how this similarity is calculated......defines the properties of face-space" (Lewis, 2004, p. 30). Valentine's (1991) exemplar-based face-space is grafted on to an Euclidean space, which characterizes similarity as a monotonic decreasing function between inter-point distances in metric space. Valentine theorized that an identity decision involves evaluating the following similarity metric: 1) the distance between the location of the stimulus and the nearest known face; 2) the distance between the stimulus and the next nearest exemplar. Hence, ease of recognition is defined by a limited subset of exemplars. This nearest neighbour algorithm pervades other incarnations of MDS, namely the Voronoi (Lewis & Johnston, 1999) and manifold (Craw, 1995) models[1].

Lewis (2004) devised a development of MDS entitled 'face-space-R'. Whilst this model utilizes Euclidean principles, ease of recognition depends on: 1) activation of the exemplar most similar to the stimulus; 2) activation of all other exemplars (proportional to exemplar density at the probe). Thus, the implementation resembles the interactive activation account of FRUs (Burton et al., 1990; 1999). A mathematical implementation of the space captured distinctiveness effects, other-race effects and the caricature advantage. Its predictive power was further bolstered by the inclusion of a 'strength parameter' which signalled familiarity. This allowed each face encountered to be encoded in space, regardless of familiarity, thus preserving Valentine's (1991) original tenet that MDS houses

[1] Craw (1995) does not explicitly reference this algorithm but the local properties of the model associated with intra-class discrimination are granted Euclidean characteristics consistent with face-space.

faces that have been seen previously but are not necessarily familiar. This is an important development since Lewis and Johnston's (1999) Voronoi model: given that identity regions saturate face-space, every encountered face could potentially activate a representation leading to false recognition. Lewis and Johnston did introduce a threshold of activation but "such an arbitrary threshold does not sit well with the explicit nature of the Voronoi model" (Lewis, 2004, p. 56).

1.3 Distinctiveness

The study of distinctiveness in the domain of face recognition has attracted a lot of attention since the early 1970s. Recognition memory tasks demonstrated that previously unfamiliar distinctive faces are more accurately recognized and less likely to be misidentified as having been seen before. A later wave of research demonstrated distinctiveness affects the speed and accuracy with which familiar faces are recognised, whether famous or personally familiar.

Section 1.3.1 reviewed the origin of distinctiveness, more specifically, whether or not it should be cast as a perceptual or mnestic phenomenon. An appreciation of the nature of this variable is an essential prerequisite to accepting Valentine's (1991) multidimensional face-space as a theoretical account of the variable. The manner in which MDS accommodates temporal dimensions of distinctiveness is a particularly pertinent issue. The relationship between distinctiveness and facial ageing (Section 1.3.2) and the developmental literature (Section 1.3.3) are considered with particular reference to MDS.

1.3.1 Distinctiveness: A Perceptual or Mnestic Variable?

A common concern uniting early studies of distinctiveness was whether or not the variable be best cast in perceptual or mnestic terms. Light, Kayra-Stuart and Hollander (1979) examined the hypothesis that distinctiveness effects arise as a result of the differential depth of processing of distinctive and typical faces with distinctive faces more attention-grabbing than typical faces. Prior to taking part in a recognition memory task, participants viewed 40 distinctive and 40 typical unfamiliar faces under incidental or intentional learning conditions (only in the latter were participants informed that their memory was going to be tested), 3 second or 8 second presentation times and likeableness or gaze direction judgment tasks.

It was supposed that the 3 second exposure, intentional learning conditions, and the gaze direction orienting task, would attenuate the distinctiveness effect by encouraging equal distribution of attention across both face types. Recognition memory performance was consistently better for distinctive faces compared with typical faces negating Light et al.'s (1979) depth of processing hypothesis. However, in another study, Hosie and Milne (1995) demonstrated that distinctive faces are maximally encoded after a very short exposure. Thus, regardless of the conceptualisation of distinctiveness as a perceptual or mnestic variable, Hosie and Milne concluded that encoding distinctive information does not require processing time.

Valentine and Bruce (1986) offered a mnestic interpretation and provided some of the first evidence that faces may be coded relative to a prototypical face, or norm. Famous faces and unfamiliar faces were rated on a 7-point scale for distinctiveness, where 1 corresponds to a 'very typical' face (famous face) or a 'very typical' face which would be 'very hard to spot in a crowd' (unfamiliar face), and 7 denotes a 'very distinctive' face (famous face) or 'very distinctive' and so 'would be relatively easy to pick out in a crowd' (unfamiliar face). Ten famous faces and 10 unfamiliar faces (half distinctive and half typical) were presented to participants in a familiarity decision task, and also in a facedness decision task (face or nonface) combined with 20 jumbled faces. In the familiarity decisions task, distinctive faces were classified faster than typical faces but in the facedness decisions task, the effect reversed. The authors hypothesised that because typical faces are relatively similar to the prototype they are classified as a face more quickly than distinctive faces, but less well recognised owing to confusion with other typical exemplars.

Whilst Valentine and Bruce (1986) concluded that distinctiveness can be accounted for "in terms of the manner in which memory traces are stored rather than the encoding processes involved in recognising faces" (p. 534), H. Ellis, Shepherd, Gibling and Shepherd (1988) subscribed to a perceptual interpretation of the facedness task. They speculated that distinctive features may trigger a face schema, regardless of the legitimacy of the stimulus, rendering classification of jumbled distinctive faces particularly difficult. Furthermore, the authors reasoned that it is untenable to couch an effect, obtained for both famous and unfamiliar faces, in mnestic terms. However, mnestic mechanisms fuelled creation of Valentine's (1991) multidimensional face-space. Moreover, recent incarnations of face-space, for example, the Voronoi model (Lewis & Johnston, 1999) and Face-space-R (Lewis, 2004) also explain distinctiveness as a mnestic variable.

1.3.2 Distinctiveness and the Ageing Face

The effects of ageing on facial appearance are wide ranging and dramatic. They include changing cephalic morphology, loss of adipose tissue, increasing wrinkles, changes in hair distribution and colour, and disproportionately large ear and nose cartilage (Alley, 1988). Yet, relatively little is known about the impact on encoding and storage of faces for recognition purposes. George and Hole (1998) demonstrated that faces are recognized despite age transformations. Participants were familiarised with six faces (with a mean age of 8.33 years old) which then had to be identified amongst a series of distractors. Photographs of the to-be-recognised faces were either identical to those learned, differed in pose and expression, or differed by age, pose and expression. Age changes comprised a 2 to 4 year increase or decrease. Recognition performance latency and accuracy for older faces resembled performance for faces differing only in pose and expression. This finding held for younger faces in the latency analysis, but performance accuracy dropped significantly below that of older faces. George and Hole concluded that representations of faces are not ageless, hence the consistently impaired performance associated with age-transformed faces. Yet neither does recognition require a perfect match of age characteristics, hence the similar consequences associated with age transformations and pose and expression changes. George and Hole entertained several possibilities for the way in which age is stored, including the suggestion that "age information is utilised in a broad way within a representation, perhaps from some kind of prototypical age face" (p.1132). Thus, it is quite conceivable that age could be coded as a dimension in face-space.

O'Toole, Vetter, Volz and Salter (1997) discovered that the application of a caricature algorithm to 3D laser-scanned heads aged the faces in addition to caricaturing them. Thus in addition to exploring an additional perceptual dimension of distinctiveness, an exploration of this relationship mandates an understanding of the way in which age information is represented. Fifty male and 50 female heads (with a mean age of 26.9 years old) were subjected to PCA represented as points located in a 99 dimensional space. The mean distance of heads from the average was 9.9. Caricature levels were set to distances from the average of 6.5 (anticaricature), 10 (estimate of veridical), 13.5 and 17 (caricatures). A subset of 60 of these faces were presented to participants at each level of caricature. Perceived face age increased linearly as a function of caricature level, to the extent that all participants perceived faces presented with the level-17 distortion as at least 50 years old.

The authors observed that caricaturing does not invoke head shape changes. Rather, facial wrinkling is amplified, and loss of adipose tissue, and skin elasticity and muscle tone under the jaw is extrapolated. O'Toole et al. (1997) reasoned that cues to ageing from the 3D representation must only have been based on the distinctive aspects of individuals relative to the adult face because no normative cues to ageing were exploited. Hence, the indirect claim that distinctiveness taps age information.

Deffenbacher, Vetter, Johanson and O'Toole (1998) explicitly investigated rated distinctiveness and memorability as a function of caricature, or perceived age, using O'Toole et al.'s (1997) 3D stimuli. Participants rated 30 of these faces for distinctiveness. Distinctiveness was related linearly to distance from the average face in 3D-grounded space. Given that age had previously varied with caricature in the same way (O'Toole et al., 1997), Deffenbacher et al. (1998) concluded that faces become more distinctive with age. Indexed by accuracy and the number trials required to learn a face, memorability increased as a function of caricature too. Thus, older distinctive faces are more memorable than younger faces located closer to the average face. That facial age taps a component of distinctiveness has important implications for studies which investigate distinctiveness with a temporal dimension. Regardless of whether or not age transformations extend vector lengths (norm-based), cause shifts within Voronoi cells (Lewis & Johnston, 1999) or create new manifolds (Craw, 1995), it is conceivable that representations change location in MDS with age. Hence, present experiments investigating distinctiveness with AoA as a temporal dimension may need to take account of ageing.

1.3.3 Distinctiveness from a Developmental Perspective

H. Ellis (1992) was the first to document the finding that distinctiveness effects only emerge beyond a certain age. Participants aged 6 to 8 years, 9 to 11 years and 12 to 14 years, took part in a recognition memory test with men's faces. Distinctiveness did not modulate performance in the youngest group. Conversely, the older groups demonstrated the customary advantage in recognising distinctive faces. From a visual inspection of the graph, however, it appears that the full advantage was not necessarily reached until 12 to 14 years of age. H. Ellis also found that for all age groups, distinctive and typical faces were equally likely to be identified erroneously as distractors (cf adult data, e.g. Bartlett, Hurry & Thorley, 1984). H. Ellis (1992) reasoned that children may fail to encode those physiognomic aspects which render a face distinctive and store typical and distinctive faces in a similar fashion to one another. Consequently, he proposed two alternative versions of face-space to accommodate these findings (see Figure 1.5). The uniform model is characterised by parameters akin to those structuring adult face-space. However, it is populated less densely and as such has an attenuated density gradient. The alternative differential model is smaller than adult face-space but preserves the differential distribution of typical and distinctive faces.

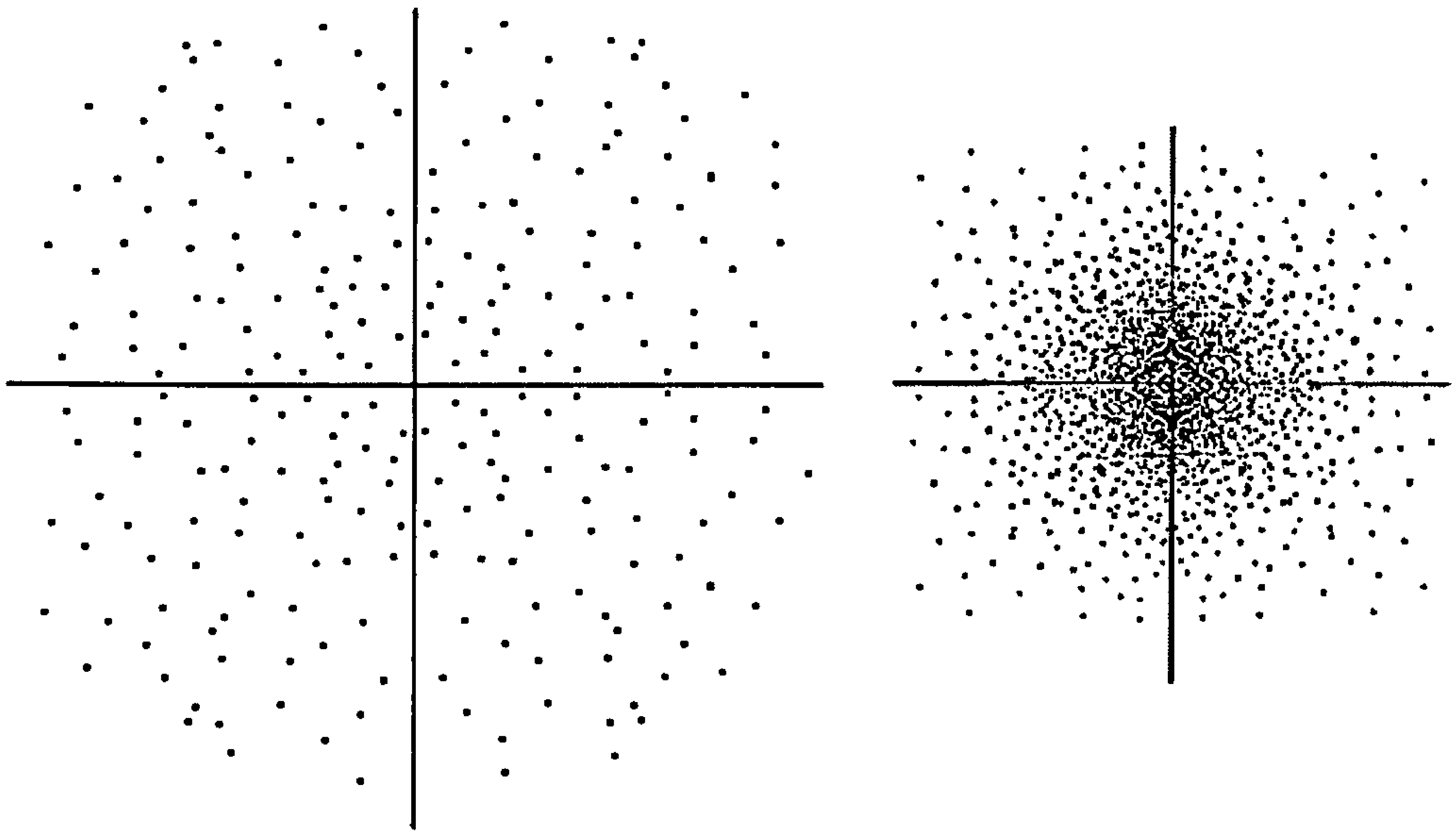


Figure 1.5. *H. Ellis's (1992) uniform model (left) and differential model (right).*

Johnston and H. Ellis (1995) sought to delineate these models and, moreover, establish whether norm-based or exemplar-based conceptualisations would provide a better account of child face-space. A recognition memory task and facedness task were administered in pursuit of this differentiation. The authors reasoned that whilst both norm and exemplar versions posit similar mechanisms to explain recognition memory performance (the distance or vector angle between target and neighbour), mechanisms dictating classification performance differ between versions. The advantage imparted to typical faces in a classification task stem from a decreased vector length between the target and prototypical face or the high exemplar density around the typical face, for norm and exemplar versions, respectively. Whereas vector length and exemplar density happen to be negatively correlated in adult space, this is not necessarily the case for child space. Hence, concomitant analyses of the two tasks could shed light upon the most apt model.

The tasks were presented to groups of participants aged 5, 7, 9, 11, 13 and 20 years old. In the recognition memory task, participants had to make old or new decisions to 18 previously studied unfamiliar males (half distinctive, half typical) intermixed with 18 distractors. Recognition performance, indexed by reaction times and d' (an integrated measure of hits and false positives) was unaffected by distinctiveness for the 5 and 7 year old groups, whereas older children and adults demonstrated the customary distinctiveness advantage. Interestingly, a visual inspection of the graph presenting d' data revealed that 9 year olds are still marginally less affected by distinctiveness than older age groups. Analyses of hits and false positives yielded distinctiveness effects across all ages and mixed results, respectively. Conversely, the classification task carried out with intact and jumbled versions of a subset of the stimuli revealed that decision latencies across all ages were affected by distinctiveness. This was mirrored by a trend in the accuracy analysis.

Johnston and H. Ellis (1995) deduced that the uniform norm-based model or differential exemplar-based model was most suitable. They reasoned that a norm-based conceptualisation must preserve the parameters and hence vector lengths of an adult model. After all, the compressed parameters and shorter vectors associated with the differential model would results in an attenuated effect of distinctiveness on classification performance. The strength of this model resides in the parsimonious way in which it can explain performance across the tasks by appealing to different mechanisms. Alternatively, the authors pointed out that an exemplar-based conceptualisation must maintain the differential density of typical and distinctive faces to cause effects of distinctiveness on classification: the attenuated density gradient of the uniform model would result in an elimination or at least attenuation of distinctiveness effects on classification. However, to allow a differential exemplar-based model to explain recognition memory performance, Johnston and H. Ellis commented that a "disentanglement of distance from target face to neighbour and the local exemplar density of target face" (p. 466) is required. Consequently, they theorized that recognition becomes impaired when the nearest neighbour is within a certain distance from the target. This way, both typical and distinctive representations have neighbours within that radius rendering recognition memory performance impaired across all representations. Further, Johnston and H. Ellis noted that if all faces are coded as 'typical' representations, this explains the poorer recognition performance associated with young children. On the other hand, the norm-based uniform model treats each face as 'distinctive' which does not sit so comfortably with performance.

Chang, Levine and Benson (2002) hypothesised that children encode faces relative to a prototypical face. Therefore, any model of child face-space should be norm-based. Applying caricature algorithms as a proxy for distinctiveness, groups of participants aged 6, 8, 10 and 27 years took part in a speeded naming and likeness judgment task. Stimuli included personally familiar faces, presented at 5 different caricature levels: -36% and -18% (anticaricatures); 0% (veridicals); 18% and 36% (caricatures). A linear relationship between naming latency and caricature level was reported with caricatured faces responded to fastest. This did not vary with participant age. In the likeness task, adults preferred veridical depictions as the best likeness of the original, whereas 6 to 10 year olds most frequently selected positive caricatures. The authors concluded that adult-like expertise is not necessary in order to make use of distinctiveness information emphasized in caricatures. Based on this premise, children as young as 6 can code whatever is distinctive about a face. Further, their susceptibility to caricature effects suggested that children are able to employ norm-based coding. Hence, the favoured encapsulation of childhood distinctiveness effects, a norm-based interpretation.

Chang et al.'s (2002) results are not necessarily in mutual exclusion with those of Johnston and H. Ellis (1995). Both face classification and caricature tasks are united insofar as they tap the facility to encode faces relative to a norm. Johnston and H. Ellis's uniform norm-based model could feasibly account for Chang et al.'s results whilst also being consistent with the lack of distinctiveness effects in their recognition memory task. It is possible that Johnston and H. Ellis's differential exemplar model could also provide a reconciliation between the two studies given that the postulated mechanisms underlying recognition memory tasks are independent to those causing classification and caricature effects. As different means of indexing distinctiveness effects subscribe to different mechanisms, it is not logical to conclude that the emergence of distinctiveness effects in one task implies the emergence of this effect in tasks which recruit different mechanisms.

Before closing this section, it is worth mentioning Valentine's (2001) observation that Lewis and Johnston's (1999) exemplar-based Voronoi model, in conjunction with Johnston and H. Ellis's (1995) uniform model, provides an alternative conceptualisation of child face-space. Given that the Voronoi model codes identities as regions which consume the entire space, the sparsely populated nature of the uniform model would result in large identity regions per face, thus providing a neat explanation for category inclusion errors, characteristic of childhood face recognition. Although Johnston and H. Ellis did not nominate the exemplar uniform model as a likely candidate, the nomination process is highly speculative without future research. However for present purposes, it is sufficient to acknowledge that mechanisms subserving recognition are not affected by distinctiveness in children.

1.4 Age of Acquisition

A persistent interest in AoA was sparked 15 to 20 years ago in the domain of word recognition. This section charts the status of AoA as a stage-dependent process to a ubiquitous by-product of the way in which information is stored and accessed in the brain.

Early studies sought to delineate effects attributable to AoA from those which arose as a result of frequency. Many studies found task dissociations between AoA and frequency, or additive effects, and following Sternberg's (1969) theory of additive logic[2], interpreted this as evidence that the variables are rooted in different origins. Section 1.4.1 reviews separate-stage accounts in the domains of word, object and face recognition. More recent single-stage accounts were less preoccupied with uncovering task dissociations, stemming from the prediction that AoA and frequency effects co-occur. These accounts are documented in Section 1.4.2 and compared in Section 1.4.3.

---

[2] Sternberg's (1969) additive-factors logic theorizes that additivity arising from RT data in factorial designs implies that the underlying mechanisms can be divided into serially-arranged operations, or stages. This model has been recently affirmed (Roberts & Sternberg, 1993; Sternberg, 1998). Bruce, Dench and Burton (1993b) provide an objective review of this method. They note that an interaction does not necessarily indicate that two variables affect the same stage of processing. However, additivity of two variables results from "independent influences on information processing" (p. 42). Moreover, as a consequence of additivity in their results, Bruce et al. (1993b) conclude that two variables act in different stages in processing.

1.4.1 Separate-Stage Accounts

1.4.1.1 Effect of AoA on Word and Object Recognition

Frequency has long been considered the prime candidate for determining lexical processing speed and accuracy. An oft cited study by Oldfield and Wingfield (1964) reported that objects with high frequency labels are named more quickly than those with low frequency labels. A development of this study was carried out by Carroll and White (1973). It considered a broader host of variables related to frequency, including AoA, and submitted results generated from 94 object drawings to a multiple regression analysis, compared with the meagre 26 in the Oldfield and Wingfield study. AoA emerged as the only significant predictor of naming latency. Carroll and White also investigated whether naming latency could be affected by cumulative frequency; a multiplicative function of AoA and frequency. However, this investigation was not fruitful. The subsequent "rejection of the original single-stage account necessitated an explanation for separate effects of frequency and AoA" (Lewis, Chadwick & H. Ellis, 2002, p. 1228). Indeed, Lachman (1973) managed to record independent effects of both AoA and frequency on object naming latency.

There are problems with the measures used in these studies. Carroll and White (1973) employed objective and rated measures of AoA. The objective measure was derived from tabulations of frequencies of words used in children's writing, thus it may have incorporated frequency. Nevertheless, Carroll and White reported a correlation of .85 between the objective and rated measures, supporting the validity of each[3]. Unfortunately the authors observed that AoA measures were significant predictors provided that they were entered into the multiple regression analysis in combination. As a result they concluded that "each is to the same degree not completely valid as a measure of age of acquisition" (p. 91). Lachman

[3] Ratings of AoA are reputed to be valid measures of the age at which words are acquired. Barry, Morrison and A. Ellis (1997) cited the following as evidence: Gilhooly and Gilhooly (1980) found a correlation of .93 between rated AoA and the rank order of words in the norms from the 'Mill Hill' vocabulary test; Morrison, Chappell and A. Ellis (1997) found a correlation of .80 between rated AoA and children's naming performance of 297 objects.

(1973) used Carroll and White's objective measure of AoA rendering the validity of their results questionable also. Moreover, rated measures of AoA and frequency were gathered from the same participants with only 24 hours between ratings. As a result, Lachman conceded that the frequency ratings used in this study may have tapped a chronological dimension.

Even with improved measures of AoA and frequency, the nature of their effect on object naming remains inconsistent. In a reanalysis of the Oldfield and Wingfield (1964) data, with frequency, AoA and word length as predictors, Morrison, A. Ellis and Quinlan (1992) found AoA to be the sole determinant of naming latency. In their own study, which included rated imageability and rated prototypicality of items as members of 'natural' and 'man-made' categories, as additional variables, only AoA and word length emerged as significant predictors in the multiple regression. In a more comprehensive study, Barry, Morrison and A. Ellis (1997) inserted the following factors into a multiple regression: Celex printed and written word frequency; AoA; imageability; name agreement; familiarity; image agreement; visual complexity and word length. Data derived from 195 images were entered into the analysis, compared with just 48 from the Morrison et al. (1992) study. Object naming latency was predicted by spoken frequency, name agreement and multiplicative term comprising AoA and spoken frequency: AoA effects emerged for low frequency words only. Despite the interaction, Barry et al. (1997) were not tempted to view AoA and frequency as operating at a single stage of processing. Whilst they concluded that both variables affect the process of activating a word's phonological form for its spoken production, they specified the locus of frequency to be at the 'lemma-to-lexeme connection' and the locus of AoA at the level of the 'lexeme' itself.

A wide variety of separate-stage accounts have been proposed, with Brown and Watson's (1987) 'phonological completeness hypothesis' one of the most influential. Brown and Watson argued that AoA effects are a symptom of the way in which phonological representations are developed. As vocabulary increases and storage space decreases, words have to be stored in an increasingly fragmented fashion. Whereas early acquired words can be stored in a holistic form, later acquired words are subject to fragmentation and require more processing time for reassembly prior to production. Monaghan and A. Ellis (2002) recovered damaging evidence against the hypothesis. They reasoned that early acquired words should be harder to dissect in a phonological segmentation task than late acquired words. On the contrary, segmentation speed was not found to be a function of AoA. Furthermore, Yamazaki, A. Ellis, Morrison and Lambon Ralph (1997) found that the age at which Japanese Kanji letters entered both spoken and written vocabularies predicted reading speed. They reasoned that there must be at least two loci for AoA effects because AoA must surely affect the quality of lexical representations in the speech output and visual input lexicons.

Morrison et al. (1992) found that the time taken to categorise objects into 'man-made' versus 'natural' categories was not predicted by AoA. In conjunction with their afore-mentioned finding (p. 41), the authors deduced that AoA affects name production rather than object recognition or comprehension. However, Brysbaert, Van Wijnendaele and De Deyne (2002) criticized the semantic task used. They asserted that it is 'suboptimal' to distribute stimuli over two different categories and then report aggregated RTs irrespective of the category. It is possible that participants may have re-defined the 'man-made' versus 'natural' task as a 'yes-no' decision task. The authors stated that stimuli in a positive category are processed differently from those in a negative category. In a more robust semantic classification task, participants had to classify words as belonging to the category 'words with definable meanings' or a category corresponding to proper nouns. Brysbaert et al. (2002) recovered AoA effects in this task concluding that the locus of AoA had been interpreted too narrowly and should be extended to encompass the semantic system[4].

1.4.1.2 Effect of AoA on Face Recognition

The *a priori* advantages of investigating AoA in the domain of face recognition are two-fold. Firstly, it is hard to disentangle mechanisms rooted in *age* of acquisition and *order* of acquisition as the two are "perfect correlates for words learnt in the

---

[4] Brysbaert et al. (2000) cite the work of Taft and van Graan (1998) as evidence that a word / first-name semantic classification task does not recruit phonology. Though regular definable words and irregular definable words exhibited a reliable difference in naming latencies, semantic classification latencies did not differ.

participant's first language" (Lewis, Chadwick & H. Ellis, 2002, p. 1230). Moore and Valentine (1999) point out that the phonological completeness hypothesis is consistent with effects arising from a critical period of language development and the proposed phonological output locus of AoA. Studies addressing effects of AoA on face processing provide a means of delineating age from order effects: the early faces used by Moore and Valentine (1998; 1999) are encountered between 6 and 12 years of age whereas early acquired words are typically encountered between the ages of 2 and 6 years. Thus, face recognition studies are a necessary prerequisite to hypothesizing mechanisms free from maturational constraints. Secondly, face recognition studies provide an important vehicle for by-passing problems associated with the high inter-correlations between AoA and other variables, such as frequency, which are more pronounced in other domains, for example word naming. Inter-correlations pose problems for interpretation of multiple regression analyses and have adverse practical implications for factorial designs.

In the late 1990s, researchers were still entrenched in the notion that AoA and frequency stem from different origins and hence, reside in different loci. Moore and Valentine (1998) investigated the effect of AoA on famous face naming latency in a multiple regression analyses based on responses to 106 faces and in a factorial analysis based on 50 critical items (25 early acquired and 25 late acquired). Participants were asked to rate when they "first became aware of each celebrity" on a 7 point scale (1 = unknown; 7 = celebrity acquired over 18 years of age). Rated distinctiveness, degree of rated familiarity with the celebrity, log surname frequency (derived from a count of surnames in the telephone directory) and number of phonemes in the full-name were entered into the regression, or controlled in the factorial analyses. The multiple regression recovered significant effects of AoA and familiarity on naming speed latency and accuracy. In the factorial analyses, early acquired faces were named faster and more accurately than late acquired faces. Consistent with the literature from other domains, AoA effects were attributed to the phonological output lexicon, the final stage in face processing (Bruce & Young, 1986; Burton et al., 1999).

There are a number of criticisms regarding the nature of the variables employed in this study. Lewis (1999a) noted that the AoA ratings may reflect a non-linear mapping of the age at which a face is acquired. He noticed that whilst points 2 to 5 on the rating scale are separated by 3 year increments, points 5 to 7 are separated by increments greater than 3 years. Although potentially problematic for a multiple regression analysis, factorial analyses should be sufficiently insensitive to this non-linear transformation of AoA ratings. Ratings of familiarity were obtained by asking participants to recall how many times the celebrity had been encountered. Moore and Valentine (1998) considered this a proxy for cumulative frequency. However, Lewis warned that recent occurrences were more likely to contribute to the estimate than earlier occurrences, hence this estimate may be distorted by current frequency. Also, the measure of surname frequency seems inappropriate on several counts: 1) experimental stimuli are full-names and full-names tend to be unique; 2) Moore and Valentine concede that "surname frequency is not analogous to word frequency" (p. 490), the frequency with which surnames are encountered has an idiosyncratic component; 3) celebrity surnames are often more distinctive, and encountered more frequently, than surnames which would appear frequently in the telephone directory. Finally, familiarity, distinctiveness and AoA were rated by all participants. However, given the overlap evinced by familiarity and AoA, the collection of ratings from a single population is dubious.

A shift towards a single-stage school of thought is largely documented by Moore and Valentine (1999). The same 50 critical items for which an effect of AoA occurred in face naming were subject to tasks involving reading aloud printed names, and face and name familiarity decision tasks. AoA was manipulated factorially and experimental sets matched for familiarity, distinctiveness, surname frequency, initial phonemic power, name letter and phoneme length. All tasks revealed AoA effects. Assuming that phonology is automatically activated by printed proper names, the reading aloud task and name familiarity decision are united in their support for a locus of AoA at the phonological output lexicon. However, phonology is not automatically activated by the presentation of a face[5], leading to the conclusion that AoA effects would have to operate at the level of the PINs (Bruce & Young, 1986) or at earlier stages, in order to influence familiarity decisions. Moore and Valentine hypothesised that multiple loci were necessary to

[5] Valentine, Hollis and Moore (1998) established that whist face naming could prime a subsequent name familiarity decision, face familiarity decisions did not prime a name familiarity decision.

account for the diversity of AoA effects or alternatively AoA reflected a "general property of the mental representation of perceptual and lexical information" (p. 420). Furthermore, they hypothesised that whilst AoA may affect the representations themselves, frequency may affect the strength of connections between representations. Still convinced by a separate-stage school of thought, the authors postulated that the challenge for future cognitive models is to account for both AoA and frequency effects.

1.4.2 Single-Stage Accounts

1.4.2.1 Cumulative Frequency Hypothesis

The cumulative frequency hypothesis, the first of the single-stage accounts, regards AoA and frequency effects to stem from the way in which information is stored and accessed in the brain. Lewis (1999a) argued that AoA is reducible to cumulative frequency, or the total number of instances of a stimulus, where this is determined by frequency of encounters per unit time (estimated by frequency) and the total time that the stimulus has been known (estimated by AoA). Whilst Carroll and White (1973) rejected a cumulative frequency explanation on account of their additive results, Lewis' revision of the hypothesis states that data needs to be log-transformed prior to analysis. Any resulting additive effects would be indicative of multiplicative terms in the instance-based model.

Lewis (1999a) tested his reformulation with a semantic classification task involving the classification of 185 famous faces as belonging to one of two soap operas. After the task, participants rated characters according to frequency of appearance on the show. Objective calculations determined how long the character appeared in the show and, in appropriate, low long since they left. A multiple regression analysis revealed that categorisation latency was significantly predicted by all three log-transformed variables. Lewis concluded that in absence of any parsimonious accounts of AoA, effects across all domains may be best explained as a simple accumulation of instances.

The task suffered from a number of short-comings predominantly highlighted by Moore, Valentine and Turner (1999). As far as stimulus selection is concerned, Lewis (1999a) included actors even if famous for other roles, provided that their role in the soap opera was their most famous. He reasoned that a semantic classification task requires forging associations between a face and its category, however, it is *a priori* likely that an actor famous for a role prior to the soap opera will be classified faster than a later acquired face, regardless of the importance of 'forged associations'. After all, Moore and Valentine (1998; 1999) demonstrated that AoA affects the processing of famous faces generally. To compound this, Moore et al. (1999) noted that Lewis included actors famous prior to their role in the soap opera. Owing to the close semantic associations between characters, Moore et al. also noted that RTs could be affected by associative priming and an accumulating semantic activation which occurs as a result of presenting a large number of celebrities from the same category. However in response to this, Lewis (1999b) stressed that stimuli were randomized, thus eliminating systematic bias conferred through priming. Another criticism regarding the design of the task stems from Brysbaert et al.'s (2000) notion that it is 'suboptimal' to distribute experimental stimuli over two categories (see p. 42). Further, Moore et al. believe that Lewis (1999a) should have controlled for familiarity and distinctiveness. In defence, Lewis (1999b) denounced familiarity as 'ill-defined' and a composite of other variables. Also, given that Moore and Valentine (1998; 1999) viewed familiarity as a proxy for cumulative frequency, it is counter-intuitive to control for a variable synonymous with the experimental variable itself. Lewis (1999b) argued that distinctiveness should be treated as a random factor and that there are no reasons to suppose that a "character's distinctiveness will influence their frequency or length of time on a TV show" (p. 312). However, Moore and Valentine (1998) did recover a correlation between AoA and distinctiveness. This correlation could reflect the inadequacy of their own rating: if early acquired faces are rated as more distinctive than late acquired faces, then the ratings may be tapping the distinctiveness of the person rather than their face. In exploring the effect of AoA on face-space, it is possible that this thesis may shed light on this dispute.

One of the most common sources of criticism of the cumulative frequency hypothesis comes from research which has found an additive relationship between

AoA and frequency. According to Sternberg's (1969) additive factors logic, additivity implies that AoA and frequency are rooted in different origins. However, Lewis, Gerhand and H. Ellis (2001) asserted that evidence against the hypothesis is flawed as a result of inappropriate analyses. Data from two influential studies, which proclaimed an independence of the two variables were re-analysed: Gerhand and Barry (1998) explored word naming latencies in a factorial design, and Carroll and White (1973), object naming latencies using multiple regression analyses. Both studies found effects of frequency and AoA, but no interaction emerged. Lewis et al. (2001) suggested that owing to the violation of the homogeneity of variance in Gerhand and Barry's data, and the presence of curvilinearity in Carroll and White's analysis, it was most appropriate to log-transform the data. Also, the cumulative frequency hypothesis suggests that this transformation is a necessary step in analysing non-linear effects. A re-analysis of both studies revealed effects of AoA and frequency, but no interaction. Additivity as a product of a transformed analysis implies multiplicity. Lewis et al. insert the caveat that these re-analyses are insufficient to suggest that AoA effects are reducible to cumulative frequency, instead the data is merely consistent with a cumulative frequency interpretation.

Moore et al. (1999) documented task dissociations to discredit the cumulative frequency hypothesis. For example, they recalled the effect of AoA on auditory lexical decision latencies in the absence of frequency (Turner, Valentine & A. Ellis, 1998) and inferred that the mechanisms which give rise to AoA effects are not those which cause frequency effects. However, Lewis (1999b) insinuated that the literature should not be selectively scrutinized to bias evidence. If one selects the Morrison et al. (1992) paper which found effects of AoA on object naming latency in the absence of frequency, for example, it can be compared with Gerhand and Barry's (1998) replication which recovered effects of AoA and frequency.

Moore et al. (1999) also rejected the cumulative frequency hypothesis on account of findings which deem effects of AoA to be task dependent: the cumulative frequency hypothesis does not specify how the effect of AoA could be influenced by task They compared the results of Moore and Valentine (1998; 1999) which found AoA effects on face naming and familiarity decisions, with those of Moore

(1998), cited in Moore et al., which recovered no AoA effect on semantic classification latency. In reply, Lewis (1999b) asserted that to negate the cumulative frequency hypothesis, Moore would have to have shown that semantic classifications were affected by frequency in the absence of AoA. As it stands, there is no evidence to suggest that AoA and frequency are dissociable. This is a surprising admission given that the strength of this hypothesis resides in its parsimony. Recall, however, that Lewis (1999a) found AoA to affect semantic classifications. Lewis (1999b) attributed the discrepancy between his results and Moore's (1998) to the differences in task requirements: classifying faces as belonging to one of two soap operas is deemed to be a natural decision compared with classifying faces according to whether they have appeared as a chat-show host (Moore, 1998), which Lewis described as an ad hoc category. Thus, once again, the cumulative frequency hypothesis is refuted on the basis of selective screening of the literature alone.

One attack, harder to counter, was the observation that frequency is dependent on stimulus modality. Turner et al. (1998) found that frequency affected visual lexical decisions but not auditory lexical decisions. Lewis (1999b) ascribed the relationship between frequency and forgetting as a means of escaping this attack; high frequency words have a recency advantage over low frequency words that is assumed to be more acute for visual stimuli than auditory stimuli. As a result, frequency effects could occur in a visual lexical decision task without recourse to frequency of occurrence as the explanatory construct. This is highly speculative and Lewis afforded no explicit reasons as to why a recency effect would vary with modality. However, to derive the bulk of criticism against the cumulative frequency hypothesis from the results of a single study is not a convincing counter-attack either.

1.4.2.2 Growing Network Model

Brysbaert et al. (2000) theorized that AoA could influence the organization of the semantic system. Given that it is a highly clustered, gradually acquired network where the meaning of late acquired concepts is linked to earlier acquired concepts, an effect of AoA is indeed mechanistically and functionally plausible. Steyvers and Tenenbaum (2005) conceptualised this suggestion as a growing network model; a single structural substrate which can account for both AoA and frequency effects. Steyvers and Tenenbaum carried out statistical analyses of semantic structures, including 'Roget's Thesaurus', and demonstrated that they posses certain structural characteristics. They reported that these small-world structures comprise a small number of well-connected nodes which form highly connected neighbourhoods. They proposed that this architecture emerges as a result of a growth process governed by a 'preferential attachment principle' which suggests that newly acquired nodes are linked preferentially to existing nodes with lots of connections. Thus, nodes which enter the network early have more connections than later acquired nodes. In addition to this 'order of acquisition' effect, Steyvers and Tenenbaum theorised frequency to be a process which modulates the probability of connections taking place. As a consequence, the size of a node's neighbourhood correlates with AoA and frequency.

The authors validated these claims by carrying out a correlational analysis of word frequency, AoA and node degree factor (a measure of centrality) in predicting naming and lexical decision latencies extracted from databases. Earlier acquired, high frequency words which are semantically more central, had the quickest RTs in both tasks. Although the model predicts combined effects of AoA and frequency in word processing tasks, the cumulative frequency hypothesis demands an equivalence of frequency and time known (AoA) in the equation that predicts performance; the growing network model makes no such claim. "The importance of AoA relative to frequency depends on the weight given to the utility parameter for determining the likelihood of connecting new nodes to particular existing nodes" (Ghyselinck, Lewis & Brysbaert, 2004, p. 49).

1.4.2.3 Neural Plasticity Hypothesis

The growing consensus that AoA effects are a fundamental by-product of the way in which information is stored and accessed in the brain was reflected by connectionist models seeking to simulate the variable (A. Ellis & Lambon Ralph, 2000; Smith, Cottrell & Anderson, 2001). A. Ellis and Lambon Ralph demonstrated that speculative analogies between AoA and catastrophic interference were misguided, owing to the misconception that old information is replaced by new information. Rather, when the training of early and late acquired items proceeds in a cumulative and interleaved fashion, akin to vocabulary development, AoA effects are an emergent property of learning.

A distributed network trained by backpropagation mapped random patterns expressed over 100 input units onto patterns expressed over 100 output units. The input patterns were binary patterns distributed over the units. Each output pattern was a copy of the input pattern except that 10% of the units had been changed from 0 to 1, or vice versa. A subset of patterns was presented at the start of training and another subset after the network had started to learn the original subset. In simulation 4, 'early' patterns were trained over 1000 epochs (where an epoch is a single presentation of input-output patterns) and then 'late' patterns were trained alongside the early patterns for a further 1000 epochs. To equate cumulative frequency, during the final 1000 epochs, the late patterns were presented twice per epoch. An analysis of the quality of the representations demonstrated an advantage for early items compared with those entered into training later, despite equating cumulative frequency.

The authors argued that early patterns exerted a greater impact on the structure of the input-output patterns than patterns encountered later, owing to a loss of plasticity. This occurred because the rate of change of the weights on connections between units reduces over the course of learning. Changes in weights are greatest at the start of training when the activation strength between input and output units is in the middle range. As activations are shifted to a more extreme value with training, successive weight changes become progressively smaller. Early items make the biggest changes to connection strengths and therefore configure the network to their preferred structure. Late items may seek to reconfigure the network but so long as early items continue to be trained, the network structure will continue to favour the early items over the late ones. Similarly to the growing network model, the neural plasticity hypothesis does not demand an equivalence of relative weights attributed to AoA and frequency. In fact, this model predicts that the weight of AoA will be significantly larger than the weight of frequency.

1.4.3 Which Single-Stage Account?

1.4.3.1 How do the Accounts Differ in the Relative Weights Attributed to AoA and Frequency?

Two recent papers explored predictions from the cumulative frequency hypothesis, the neural plasticity hypothesis and the growing network model concerning the relative weights ascribed to AoA and frequency in their respective implementations. Lewis et al. (2002) ran a set of neural network simulations in parallel with a behavioural experiment. The experiment comprised a semantic classification task akin to that employed by Lewis (1999a). Two sets of participants, one set with an average age of 20 years, and the other with an average age of 60 years were recruited. The model was trained to represent younger and older participants; to capture older participants, the network learned background items prior to training. Indexed by classification latency and network error, the respective behavioural task and simulations yielded significant coefficients of AoA and frequency for both populations. Furthermore, the simulation revealed that each of the AoA coefficients were consistently larger than the frequency coefficients. This same pattern emerged from an analysis of the older participants. Lewis et al. concluded that however one explains the effects found with the younger participants, results from the network and older participants support the tenet of the neural plasticity hypothesis which states that AoA effects cannot be recast as cumulative frequency effects.

These findings were reinforced by Ghyselinck, Lewis and Brysbaert (2004) who conducted a systematic comparison of AoA and frequency coefficients. Eight behavioural tasks, which differed in their reliance on information such as perceptual information, phonology, articulation processes and semantic information, were administered. A strong linear relationship between the magnitude of AoA effects and the magnitude of frequency effects emerged across the tasks. Moreover, the weight of the AoA parameter was consistently larger than that of frequency. Once again, these findings afford support for the neural plasticity hypothesis. The authors also noted that this pattern of coefficients is consistent with Steyvers and Tenenbaum's (2005) growing network model.

1.4.3.2 Staged Inputs: An Imposition or Simulation of Reality?

Whilst A. Ellis and Lambon Ralph (2000) and Steyvers and Tenenbaum (2005) purported that inputs should be staged in order to mimic the gradual acquisition of vocabulary or concepts, other studies which have simulated AoA effects argue that staging inputs is not necessary (Smith, Cottrell & Anderson, 2001; Zevin & Seidenberg, 2002). Smith et al. defined AoA as the point at which error in the network falls below a given threshold and thus, early acquired words are those which are easiest to learn. Zevin and Seidenberg posited that AoA may be determined by frequency of occurrence and should be considered as part of a theory which addresses why certain items are learned earlier than others. Hence, implementation of staged inputs would be considered an imposition which negates exploration of factors that determine AoA.

Working on the basis that words learned more frequently early in development will be learned earlier than words mainly trained later, Zevin and Seidenberg (2002) carried out a distributed connectionist network simulation to test the hypothesis that the frequency trajectory of experience with a word has an effect on ultimate network performance independent of cumulative frequency. 'Early acquired words' were trained more frequently early in training; they were presented to the network at a frequency of 1000 for the first 3 epochs, 500 for the next 4 epochs, the following epochs received a frequency of 100, 50 and 10 in succession, culminating in a frequency of 1 for the final 3 epochs. The 'late words' were characterized by a complementary set of trajectories. Whilst being trained with the 'early' and 'late' items, the training corpus also included a number of background items. When cumulative frequency was matched upon termination of training, no reliable differences in network error were found between the early and late sets. The authors observed that an advantage was imparted to early words early in training, but this was simply a frequency effect.

A. Ellis and Lambon Ralph (2000) simulated vocabulary acquisition, Steyvers and Tenenbaum (2005), the age of acquisition of concepts and as Zevin and Seidenberg (2002) pointed out, their research is concerned exclusively with the acquisition of orthographic-phonological correspondences. This knowledge is not applied when acquiring a spoken vocabulary, acquiring concepts, nor does it bear any relevance to face processing. Thus, the learning principles dependent on frequency of presentation, associated with Zevin and Seidenberg's model of reading, may not generalize to other tasks.

1.4.3.3 Conditions Yielding AoA Effects: A Reconciliation of Models

Zevin and Seidenberg (2002) carried out a series of simulations to probe conditions susceptible to AoA effects. Unlike their first simulation (described on p. 52) which incorporated words with consistent spelling-sound correspondences, the early and late items entered into the third simulation had minimal orthographical-phonological overlap and the two sets themselves also had minimal overlap. With frequency trajectories similar to Simulation 1 and cumulative frequency equated, a significant advantage was found for words presented frequently early in training, compared with those with the reverse trajectory. Similarly, Monaghan and A. Ellis (2002) reported effects of AoA in a word naming task and counterpart simulation where words had inconsistent mappings between orthography and phonology. Moreover, AoA effects were attenuated when input and output patterns were predictable.

Zevin and Seidenberg (2002) argued, however, that conditions which gave rise to AoA effects in these simulations (A. Ellis & Lambon Ralph, 2000; Monaghan & A. Ellis, 2002; Zevin & Seidenberg, 2002, Simulation 3) are not characteristic of reading an alphabetic orthography. The authors noted that the patterns learned in the simulations by A. Ellis and Lambon Ralph and Monaghan and A. Ellis, did not possess the statistical regularities of an alphabetic orthography. As such they posited that the network could only learn by a process of pattern memorisation to overcome this degree of arbitrariness. However, these simulations had not been intended to model reading or lexical processing. Also Zevin and Seidenberg's simulations are not without their problems. In the first two simulations, those which failed to show AoA effects, 96.3% of the training corpus constituted background items, with only 3.7% of the corpus early or late items. A constant training of background items is not psychologically plausible. Moreover, these items may have subdued AoA effects in these simulations by aiding the assimilation of early or late items into the network. It may be no coincidence that when the background items were dispensed with in simulation 3, AoA effects emerged.

Nonetheless, there does seem to be consensus that AoA effects are present when arbitrary or inconsistent mappings characterise the relationship between input and output patterns. As Zevin and Seidenberg (2002) put it, "one task that probably yields genuine AoA effects is learning the names associated with faces" (p. 29) or indeed any face recognition task. The principles generated by the simulations (A. Ellis & Lambon Ralph, 2000; Monaghan & A. Ellis, 2002; Zevin & Seidenberg, 2002, Simulation 3) remain unfaltered for the present concerns.

1.5 Relationship Between Identity and Gender Processing

Historically, identity and gender have been conceived of as independent processes, hence their functional separation in the Bruce and Young (1986) model. More recently, less radical versions of this parallel-route account, and even single-route conceptualisations, have pervaded the literature seeking to explain the seemingly interdependent relationship between the dimensions.

This section opens with a review of the physical bases of identity and gender, establishing that there is a perceptual overlap between the dimensions (Section 1.5.1). Section 1.5.2 tracks the shift from parallel-route to single-route conceptualisations, outlining a common trait of the literature at present; the whimsical way in which single-route accounts and less radical formulations of the parallel-route account are selected to explain the relationship. Studies which offer evidence that multidimensional face-space codes gender are reviewed for their value in opening up the possibility that MDS serves as an interface between identity and gender processing (Section 1.5.3).

1.5.1 What are the Physical Bases of Identity and Gender?

1.5.1.1 Physical Basis of Identity

Over the course of familiarisation, internal features (eyes, nose, mouth) become increasingly important cues for identification compared with external features (hair, face shape, chin), the salience of which don't tend to vary as a function of face familiarity. Indeed, this has been shown in a famous face identification task, recognition memory task with famous and unfamiliar faces (H. Ellis, Shepherd & Davies, 1979) and a whole to internal or external face matching task with famous and unfamiliar faces (Young, Hay, McWeeny, Flude & A. Ellis, 1985). Moreover, Clutterbuck and Johnston (2002) demonstrated that this qualitative processing shift is a graded function of familiarity. In a whole to internal or external matching task, the internal features of highly familiar celebrities were matched significantly faster than those of the moderately famous, which in turn, were matched significantly faster than unfamiliar internals. Once again, external feature matching did not vary with familiarity. Similarly, Bonner, Burton and Bruce (2003) discovered that internal feature matching of faces which had been familiarised over 3 days improved relative to the processing of both unfamiliar internals and familiarised or unfamiliar externals.

Notwithstanding external features, which do contain information important for identification, this body of literature suggests that internal features are vital in the establishment of representations which mediate identification: whereas familiar face processing receives information from internal and external features, unfamiliar faces rely predominantly on external features[6]. Taking this one step further, Young, Hellawell and Hay (1987) demonstrated that the configurational processing of internal and external parts of famous faces were equally disrupted by pairing the internal parts of famous faces with the externals from others. Thus, configurational information is important in identifying faces whether from internals or externals.

[6] With exception to H. Ellis, et al. (1979) who found that familiar faces rely predominantly on internal features, whereas an equivalence of information from internals and externals characterises unfamiliar face processing.

Bruce and Langton (1994) demonstrated the importance of configurational information and pigmentation in enabling a face to be recognised. The authors reported independent effects of inversion and image negation on the time taken to identify famous faces, however only inversion affected laser-scanned images. It was concluded that inversion disrupted the processing of configurational information present in both stimulus types and negation disrupted recognition by inverting pigmentation values, hence why the laser-scanned images, devoid of pigmentation, remained unaffected by negation.

Individual physiognomic features have been found to be particularly important for identification. Roberts and Bruce (1988) found that familiarity decision latencies were significantly disrupted by masking the eyes of famous and unfamiliar faces, whereas obscuring the nose or mouth exerted little effect. A similar conclusion was reached by O'Donnell and Bruce (2001). Seven familiarised faces and 7 unfamiliar faces were presented in a same-different decision task with internal (eyes, mouth) and external (chin, hair) feature manipulations. Whereas changes to hair were the sole unfamiliar face manipulations detected, hair and eye manipulations were observed when familiar faces were presented. The authors concluded that it is the eyes alone that benefit from the shift towards internal feature processing.

More recently, Sadr, Jarudi and Sinha (2003) investigated the importance of eyebrows for identification. Participants had to identify, either by name or semantic information, celebrities with their eyes or eyebrows digitally removed. Whilst the absence of either feature caused a significant drop in performance accuracy, recognition performance for faces without eyebrows was significantly worse than performance without eyes. The authors reasoned that the salience of this feature may stem from a bias in attention resulting from their key role in expression analysis and their status as a sexually dimorphic feature.

1.5.1.2 Physical Basis of Gender

Whereas masking the eyes considerably slowed familiarity decisions, Roberts and Bruce (1988) found that gender decisions were slowed by masking the nose. They also discovered that when presented as isolated features, the eyes provided the most reliable information for gender decisions, with the mouth and then nose resulting in increasingly slower latencies. From this, Roberts and Bruce deduced that the relationship between the nose and the rest of the face is important in determining gender, rather than the shape of the nose itself.

Conversely, in a similar paradigm, which investigated masking eye, nose and chin regions, Bruce et al. (1993a) reported a tendency for mask type to interact with face gender, insofar as masking the nose had a greater impact on the classification latency of male faces compared with females, and masking the eyes (including the eye brows) had a greater effect on female face classification latency. Brown and Perrett (1993) recruited a less restricted set of features to investigate the classification of isolated features and faces whose features had been substituted with features from a face of the opposite sex. Manipulated features comprised eyes, nose, mouth, jaw, chin and combinations of these features; brow and eyes, eyes and nose, and nose and mouth. Where features were substituted across faces, alterations to the jaw and to the brow and eye region changed the gender of the perceived face most effectively. When features were seen in isolation, brow and eyes, and brows alone, were classified according to gender most accurately. In fact, the nose was the only feature not accurately classified. Whilst there was a degree of consistency in brow and eye saliency, Brown and Perrett (1993) observed that different features affected gender classifications depending on whether they were seen in isolation or as part of the configuration formed when seen in the context of the whole face. Moreover, Bruce et al. (1993a) stated that "individual local cues to sex are perhaps not very useful compared with relationships between a multiplicity of local cues" (p. 141), hence, providing one account for the largely inconsistent findings.

As a means of investigating multiple classes of gender information, Bruce et al. (1993a) presented hair concealed male and female faces as photographic positives and negatives, with viewpoint upright and inverted. Negation was manipulated to assess the contribution of shape from shading to gender decisions and inversion was manipulated to test whether or not configurational information supports these decisions. Additive effects of negation and inversion on classification accuracy were recovered, leading the authors to conclude that configurational information and 3D information facilitate gender processing. Furthermore, gender classification performance remained above chance despite negation and inversion, confirming the importance of local features as cues to gender. Consequently, Bruce et al. concluded that three broad classes of information are used to determine gender: local cues, configurational information and 3D information.

Bruce and Langton (1994) also provided evidence for the latter information sources in gender processing. The authors found effects of inversion and negation on classification latencies. Taken together with their recognition task, Bruce and Langton were able to conclude that configurational information is utilised in identity and gender processing. They hypothesised that negation only affects the former insofar as pigmentation values are inverted, as opposed to the disruption of shape from shading proposed to underpin gender processing. The differential role of 3D information in identity and gender processing is also captured by the three-quarter view advantage known to characterise gender decisions (Bruce et al., 1993a) and the absence of this advantage in familiarity decisions (Bruce, Valentine & Baddeley, 1987b).

Burton, Bruce and Dench (1993) attempted to identify dimensions on which male and female faces differ and combine these dimensions into a single function to discriminate male and female faces. When hair is concealed, participants make gender decisions with 96% accuracy (Bruce et al., 1993a). Thus, a set of measures for discriminating faces approximating this level of accuracy was sought. Discriminant functional analyses (DFA) with Euclidean distances, complex ratio and angle measurements to full faces and profiles, and 3D measures derived from full faces and profiles failed to discriminate genders with satisfactory degrees of accuracy. When the 16 most powerful predictors, collated across the different measures, were entered into a DFA, performance accuracy reached 94%, approximating human performance levels. The authors noted that of the 5 strongest predictors, 3 of these were 3D measures which appeared to represent length of cheeks and nose protuberance (larger in males) and cheek protuberance (larger in females). As pointed out by Burton et al., these concur with Enlow's (1982) observations that males have a larger nose and more protuberant brow than females. Additionally, the relative size of these features make female cheekbones look more prominent than male cheekbones, coupled with the fact that female cheekbones are covered in a pad of adipose tissue. The remaining 2 predictors were mouth shape and eyebrow thickness, which, as the authors suggested, can be gleaned from the 2D picture plane. Like Bruce et al. (1993a), Burton et al. (1993) concluded that 3D shape and 2D information are recruited to enable gender classifications.

Given the diversity of findings, it may seem tenuous to draw parallels between the different types of information used in processing identity and gender. Nevertheless, several commonalities can be extracted. Both processes use local and configurational information. Of this information available in the picture plane, the eyes and/or brows were found to be particularly salient for identification (O'Donnell & Bruce, 2001; Roberts & Bruce, 1988; Sadr, Jarudi & Sinha, 2003), and, in addition to other features, gender processing (Brown & Perrett, 1993; Bruce et al., 1993a; Burton et al., 1993). However, specific comparisons across identity and gender processing studies are limited owing to differences between studies in the features investigated and the possible confound between task and face familiarity. In Roberts and Bruce's (1988) study where familiarity and feature manipulation was constant across tasks, the salience of the nose, and its relationship with facial structure in gender processing, became apparent. This finding is indicative of the role of facial protuberance and 3D structure in gender processing, echoed by the literature (Bruce et al., 1993a; Bruce & Langton, 1994; Burton et al., 1993; Enlow, 1982), notwithstanding its lack of salience for identification (Bruce & Langton, 1994; Bruce et al., 1987b).

1.5.2 *Accounts of the Relationship Between Identity and Gender Processing*

1.5.2.1 Independence of Identity and Gender Processing: The Parallel-Route Hypothesis

Bruce (1986) documented evidence which supported the parallel processing of identity and gender, thus preserving the architecture of the Bruce and Young (1986) model of face recognition. Gender classification performance by participants personally familiar with half of the faces and unfamiliar with the rest, was compared against performance by participants unfamiliar with both sets; this provided a baseline condition against which to assess effects of familiarity on classification latency. Although male and female faces were presented to fulfil task requirements, the 16 critical items comprised of solely male faces. The participants familiar with half of the faces categorized these faces faster than the unfamiliar faces, though this difference was not significant. Interestingly, the group of participants unfamiliar with both sets classified the 'familiar' set of faces significantly slower than the unfamiliar set. Bruce interpreted this latter finding as evidence that the 'familiar' faces were more difficult to categorize according to gender than the unfamiliar faces. As such, familiarity with faces did indeed facilitate gender decisions insofar as it worked against this difficulty in classification to speed up decisions to familiar faces relative to unfamiliar faces. However, a more compelling conclusion would have demanded that faces be explicitly matched on ratings of masculinity.

Bruce (1986) noted that this interaction between familiarity and participant group disappeared in a by-items analysis. Upon further inspection, the pivotal by-subject finding was attributed to effects carried by two 'familiar' items which were particularly hard to categorize according to gender for the participants unfamiliar with them. This sparked the tentative conclusion that the cognitive system may direct gender-relevant semantic information to facilitate gender decisions when the analysis of sexually dimorphic features is slowed. Bruce postulated that when separate decision making processes use information relevant to each other, the apparent influence of one route upon the other may emerge, provided that the time course of the routes are comparable.

Further evidence in favour of a parallel-route conceptualisation of identity and gender processing was provided by Bruce et al. (1987a) who manipulated familiarity and masculinity factorially and compared the results of a familiarity decision task with a gender classification task using the same stimuli; familiar high and low masculinity male faces and unfamiliar high and low masculinity faces matched for familiarity and masculinity. This not only provided a means of assessing the impact of familiarity on gender processing but also the influence of gender on identity processing. An analysis of the familiarity decision latencies revealed that familiar faces were responded to significantly faster than unfamiliar faces, but no effect of masculinity emerged. The corresponding gender decision analysis revealed that highly masculine faces were classified significantly faster than faces low in masculinity, but unlike the findings of Bruce (1986), familiarity with the face exerted no influence. The authors noted that gender judgments were made sufficiently rapidly (532 to 694 ms) that semantic information could not support processing: familiarity decisions ranged from 883 ms to 1113 ms. Interestingly, the two items which were hard to classify according to gender in the Bruce (1986) study were described as having RTs 'over 750 ms' yet this also seems sufficiently quick a response to negate semantic support. Nonetheless, both studies argue in favour of the parallel processing of identity and gender.

More recently, Wild et al. (2000) offered support for this hypothesis. This stemmed from the simple observation that participants classified adult faces according to gender significantly more accurately than those of 7 to 8 year olds and yet a recognition memory task yielded no difference in recognition accuracy between the two face types. The authors concluded that the greater saliency of sex signalled by adult faces compared with child faces did not improve recognition of the adult faces and therefore, gender classification accuracy is not related to recognition accuracy. Wild et al. argued that this evidence illustrates a degree of independence between recognition and classification. However, unlike Bruce (1986) and Bruce et al. (1987a) who explored the impact of familiarity on gender classifications, Wild et al.'s findings have no bearing on this relationship as both face types were equally recognisable. Rather, their findings negate an asymmetric relationship between identity and gender processing where the latter can influence the former.

1.5.2.2 Integrality of Identity and Gender Processing: The Single-Route Hypothesis

Goshen-Gottstein and Ganel (2000) and Ganel and Goshen-Gottstein (2002) pioneered the hypothesis that identity and gender are processed by a single route. In their first paper, they explored the asymmetric influence of familiarity on gender processing, by exploiting the paradigm, repetition priming on to a sex decision. Proponents of the parallel-route hypothesis attribute repetition priming effects to the strengthening of connections between FRUs and PINs (Burton et al., 1990). As a consequence, repetition priming effects are not expected to affect unfamiliar faces nor emerge during a gender classification task. Akin to superseded accounts of repetition priming (Bruce & Young, 1986), Goshen-Gottstein and Ganel suggested that this phenomenon results from the reactivation of representations stored in FRUs. Further, the authors proposed that FRUs sub-serve gender classification tasks and a single exposure of a face is sufficient for their formation. Hence, they predicted repetition priming onto a gender decision for both familiar and unfamiliar faces. However, the authors hypothesised that this effect is contingent on the nature of the to-be-processed stimulus. They argued that if participants adopt superficial hair-style heuristics, the perceptual whole is truncated and attention is directed away from internal features critical for identification. Consequently, processing will be inadequate for the formation of a representation to sub-serve priming, or in the case of familiar faces, the recruitment of existing representations (FRUs),

Goshen-Gottstein and Ganel (2000) investigated repetition priming onto gender decisions with complete and 'hair-deleted' faces. During a study phase, participants rated 56 unfamiliar faces for intelligence. Half of the participants saw complete faces and the rest, hair-deleted faces. At test gender decisions were made to hair-deleted versions of all of those faces seen at test (Experiment 1) or complete versions (Experiment 2). Regardless of face type at study, gender decisions to hair-deleted faces were primed, whereas complete faces were not. Similarly structured tasks with famous faces delivered the same results. The authors treated this finding as compelling evidence for the integrality of identity and gender processing.

As an additional comparison to help determine the validity of the parallel-route hypothesis, Goshen-Gottstein and Ganel (2000) investigated whether or not gender decisions were affected by familiarity. The rationale being that if familiarity and gender processing occur at a single locus, unfamiliar faces will command slower gender judgments than familiar faces because they have no stored representations to facilitate this decision. The gender classification latency of unstudied complete and hair-deleted familiar faces was compared with the respective unfamiliar faces. Responses to familiar faces were consistently faster than responses to unfamiliar faces, bolstering support for a single-route hypothesis.

A number of criticisms of this study are considered presently. Regarding the post-hoc analysis, Goshen-Gottstein and Ganel (2000) inserted the caveat that the different processing times associated with the familiar and unfamiliar faces "may reflect difference in the ease with which participants made sex judgments to photographs from the two stimulus sets" (p. 1208). Also, the authors offered no explanation why the familiarity effect was the same magnitude across complete and hair-deleted stimuli. Based on the premise that FRUs are not recruited to mediate gender decisions when hair-style heuristics abound, an attenuated familiarity effect for complete faces would have been a reasonable assumption: recall that Bruce (1986) found no effect of familiarity on gender decisions (except for two items) when hair-style was not conservatively edited. Regarding the two experiments, Goshen-Gottstein and Ganel reinstated the theory that repetition priming results from reactivations of FRUs rather than adopting the more recent mechanism. Without recourse to the former, Goshen-Gottstein and Ganel seem unable to explain why unfamiliar faces were primed. Thus, the single-route hypothesis is at odds with contemporary models of face recognition and lacking a coherent explanatory construct.

Whereas Goshen-Gottstein and Ganel (2000) investigated the relationship between identity and gender at relatively late processing stages, Ganel and Goshen-

Gottstein (2002) employed Garner's speeded identification task as a rigorous test of the perceptual separability between the two dimensions. More specifically, this paradigm tested whether or not attention could be successfully directed to one dimension whilst ignoring the other. Participants performed familiarity decisions and gender decisions to 40 faces (familiar males, familiar females, unfamiliar males, unfamiliar females) presented a number of times over a series of 'baseline' blocks and 'filtering' blocks. In the baseline blocks participants judged one dimension whilst the other was held constant and in the filtering blocks participants judged one dimension whilst the irrelevant dimension varied. If performance in the filtering block is worse than in the baseline block, this indicates a failure to attend to the dimensions selectively, hence Garner Interference (GI). The experiment was administered once with complete faces and once with hair-deleted faces.

In the complete face experiment, performance did not vary between baseline and filtering blocks. This spurred the conclusion that identity and gender are falsely separable when hair-included faces allow gender decisions to be made on the basis of hair-style heuristics. Conversely, GI emerged when hair-deleted stimuli were used, indicating a perceptual integrality of dimensions when attention is directed to internal features. In conjunction with their earlier study, the authors postulated that identity and gender are accommodated by a single processing route at both late (mnemonic) and early (perceptual) stages. Taken together with neuroanatomical findings (Dubois et al., 1999; Rossion et al, 2001) the authors pointed out that this single route is maintained both functionally and at a neuroanatomical level. Moreover, Ganel and Goshen-Gottstein (2002) speculated that the gender of a face is an emergent property of its identity, echoing Haxby et al.'s (2000) neuroanatomical model.

Ganel and Goshen-Gottstein (2002) observed, however, that familiarity and gender were not equally discriminable. An analysis of the latencies in the baseline blocks revealed that gender decisions were performed faster than familiarity decisions, thus interference from gender to identity could be an artefact of the ease of gender discriminability. However, the authors reasoned that because discriminability differences in the complete faces experiment caused no interference from gender to identity, any GI in the hair-deleted experiment was a genuine effect. However, this deduction is not satisfactory given that the mechanisms underlying gender judgments are different from one experiment to the other, hence what may affect one experiment may not affect the other. Even if the asymmetric GI from identity to gender judgments is the sole robust conclusion to be extracted from this study, it is nonetheless difficult to reconcile with a strict parallel-route hypothesis.

1.5.2.3 A Less Radical Version of the Parallel-Route Hypothesis

Findings which are suggestive of a single processing route can also be explained by less orthodox versions of the parallel-route hypothesis. Even Ganel and Goshen-Gottstein (2002) admitted that their data disprove only the most radical versions of the parallel-route hypothesis: two routes may exist but only if highly interconnected. Rossion (2002) found an asymmetric effect of familiarity on gender processing and Baudouin and Tiberghien (2002), the reverse relationship. Whilst Rossion did offer a single route interpretation, both studies attributed their interactive effects to a modified version of the parallel-route hypotheses.

Rossion (2002) carried out a gender classification task on participants familiarised with 30 faces and participants unfamiliar with these items. The 30 'familiarised' faces were morphed with 30 unfamiliar faces to produce 6 versions of each familiarised-unfamiliar pair, saturated at different percentages of familiarity, varying from a 0% familiarised face to a 100% familiarised face, separated by 20% increments. A latency analysis revealed an interaction between participant group (familiarised or unfamiliar) and level of saturation. Familiarised participants classified faces perceived as familiar (60% to 100% saturation) faster than faces perceived as unfamiliar (0% to 40%), whereas the other participants responded evenly across all faces.

Familiarity was concluded to facilitate gender processing[7]. Furthermore, Rossion (2002) reasoned that this was not dependent on the visual similarity between encoded and stored representations, but instead, the perception of familiarity: faces

[7] Rossion (2002) suggested that this study may have recovered this effect where other studies have failed to (Bruce, 1986; Bruce et al., 1987) because the faces were completely cropped ensuring that no gender cues were present. As the authors note, Bruce (1986) minimized external gender cues, but hair-style heuristics were still available for gender processing negating reliance on internal features.

saturated at 40% were not classified faster than faces saturated at 20%. He deduced that the cognitive system must have discriminated familiar and unfamiliar faces prior to the occurrence of gender decisions. Rossion theorized that the implementation of a cascade mechanism in the parallel-route model could accommodate the results insofar as later processes may be initiated before earlier ones are completed. Thus, identity and gender may be processed independently but identity facilitates gender decisions despite the faster processing of gender information.

Rossion (2002) offered a single-stage interpretation too. He postulated that the perceptual representations extracted during face processing are overlapping for identity and gender judgments. The finding that configurational information is important for identification (e.g. Young et al., 1987) and gender judgments (e.g. Brown & Perrett, 1993; Bruce et al., 1993a; Burton et al., 1993) permits this contention. As does the proposal that invariant properties of a face (identity and gender) are accommodated by a common neural substrate (Haxby et al., 2000).

Baudouin and Tiberghien (2002) demonstrated an effect of gender on face identification. Just as Rossion (2002) hypothesised that the cognitive system can aid gender decisions with familiarity information, Baudouin and Tiberghien hypothesised that the cognitive system can facilitate familiarity decisions with gender information. Participants had to search for an androgynous face, labelled arbitrarily with a male or female name, amongst a series of sequentially presented male and female faces. Six androgynous faces were created by morphing 6 pairs of male and female faces and these were presented in the context of 48 distractors. Participants had to recognise the target face labelled 'John' in one block, and labelled 'Mary' in another. The authors found that a face is rejected faster when its gender differs from the gender of the target face.

Baudouin and Tiberghien (2002) hypothesised that the cognitive system pre-activates facial representations and visually-derived semantic information, including gender. If the gender of the perceived face is counter to expectation, the cognitive system rejects a face without recourse to identification. The authors proposed that Valentine's (1991) MDS provides a useful framework within which to house this mechanism. Moreover, if gender corresponds to a region in MDS or a dimension in space, the process of gender categorisation may reduce the space to an area corresponding to one gender. When the gender of a perceived face is counter to expectation, the cognitive system can reject the face without having to probe a more precise location in MDS.

1.5.2.4 Reconciliation with the Parallel-Route Hypothesis

Like Goshen-Gottstein and Ganel (2000), McNeill, Burton and A. Ellis (2003) investigated the relationship between identity and gender processing in the context of repetition priming. Consistent with proponents of recent face recognition models, McNeill et al. (2003) maintained that repetition priming operates exclusively in the identity system and reflects changes due to the repeated activation of the connection between FRUs and PINs. A lack of priming onto a gender decision under normal conditions was argued to arise if participants processed superficial characteristics, rather than accessed semantics. Indeed, McNeill et al., Bruce (1986) and Goshen-Gottstein and Ganel concur in their proposition that the absence or presence of priming reflects the locus in the system where the classification is made. However, whereas McNeill et al. and Bruce contended that the presence of priming during this task reflects semantic input, Goshen-Gottstein and Ganel argued that the recruitment of FRUs and their repeated activation is responsible.

Participants in the McNeill et al. (2003) study made speeded British/American decisions to 24 famous faces, followed by a gender classification to the faces, full-names or surnames of these celebrities plus 24 novel celebrities. Only surnames were susceptible to repetition priming. McNeill et al. reasoned that when presented with a surname, devoid of superficial gender cues, participants must access semantics to decide whether a celebrity is male or female. They argued that parallel-route models can accommodate effects of identity on gender processing when gender is accessed as a semantic property. In addition, the experiment demonstrated cross-domain repetition priming onto a gender decision. McNeill et al. argued that because Goshen-Gottstein and Ganel's (2000) model relies on domain-specific perceptual records held in FRUs, cross-domain priming questions the validity of their proposal.

Clutterbuck and Johnston (2004) also discovered an asymmetric relationship between familiarity and gender processing. Like McNeill et al. (2003), they proposed a theory compatible with parallel-route models. The study employed a set of 12 famous faces and two sets of 12 unfamiliar faces. The sets were matched on ratings of distinctiveness and masculinity, and the two unfamiliar sets were also matched for similarity. Initially, participants were familiarised with one of the unfamiliar sets by making honesty ratings to full-face versions of each face: the faces were presented 10 times, each presentation lasted 2 seconds. Participants then performed gender judgments to three-quarter-view versions of all 36 faces. The external features were removed from these test items, leaving only internal features (eyes, nose, mouth), to ensure rigorous elimination of superficial gender cues. Clutterbuck and Johnston (2004) reported a graded effect of familiarity on gender decisions, with response latencies significantly increasing from famous faces to familiarised faces to unfamiliar faces.

The present finding supports previous studies which have demonstrated that familiarity facilitates gender processing when visual analysis is laborious (Bruce, 1986; Ganel & Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000; McNeill et al., 2003; Rossion, 2002). However, Clutterbuck and Johnston (2004) consider that their results are not consistent with those studies which argue for a retrieval of semantic information to aid gender decisions. They speculated that the speed with which gender decisions are performed makes timely semantic access unlikely. Instead Clutterbuck and Johnston proposed an alternative parallel-route account for the interdependence of the two variables. In a face recognition model, feature units pass activation to FRUs and units which mediate gender processing; these sets of units are activated in parallel. The authors hypothesise that FRUs pass feedback to the feature units to stabilise the representation. Consequently, gender decisions are facilitated by facial structure information plus feedback from the FRUs.

This account bypasses the need to recruit the semantic system and problems associated with the time ranges inherent in this mechanism. It also explains repetition priming onto gender decisions for unfamiliar faces, without needing to appeal to Goshen-Gottstein and Ganel's (2000) single-route hypothesis and its conditions i.e. FRUs mediate gender processing. Clutterbuck and Johnston (2004) argued that the facilitation in performance to newly familiarised faces stems from some degree of FRU formation following the familiarisation process. The authors speculated that FRU formation may be partial, owing to the restricted exposure of newly familiarised faces. But this partial formation explains why newly familiarised faces facilitate gender processing performance to a lesser degree than famous faces. This account also explains why most of the afore reviewed studies have reported asymmetric effects of identity onto gender processing: Clutterbuck and Johnston's proposed mechanism only operates in one direction. After all "it is difficult to predict any circumstances where information from the gender process could ever assist in determining the identity of the face" (Clutterbuck & Johnston, 2004, p. 166, but cf Baudouin & Tiberghien, 2002).

1.5.3 Multidimensional Face-Space: A Gender Construct

Intuitively, one would expect that MDS is characterised by relatively disparate pockets of male and female representations; dimensions coding sexually dimorphic features or configurations will inherently distil representations into dimorphic regions. Recall Baudouin and Tiberghien (2002) who hypothesised that MDS may be capable of coding gender as separate subspaces or as a dimension in itself. Moreover, the results of their study (see p. 66) led Baudouin and Tiberghien (2002) to predict that gender categorisation helps locate a perceived face in space by screening off areas containing faces of the opposite gender. Two papers which explore the feasibility of MDS as an interface between identity and gender processing are reviewed.

Johnston, Kanazawa, Kato and Oda (1997) investigated how gender may be plotted in MDS. Given the increased differentiation between male and female faces with age, gender and age were explored concurrently. In their first experiment, participants had to make age decisions (young or old) and gender decisions to 48 Japanese male and female faces. Half of the faces were of 20 year olds and half, 6 year olds. An analysis of the age judgments revealed that female adult faces were classified more slowly than male adult faces and child faces; the latter did not differ in performance. An analysis of gender decisions demonstrated the slower processing of child faces compared with adult faces.

Johnston et al. (1997) reasoned that age and gender judgments are made in a similar manner to facedness decisions (Valentine, 1991). The stimulus is encoded and compared to nearby representations. However, some of the representations will be of the same category as the perceived face whereas others will be of the opposite category thus disconfirming the judgment. Decision latencies increase with evaluation of these contradictory pieces of information. The authors reasoned that female adults faces must be located between child faces and male adult faces because age judgments to these faces were particularly slow. Further, because child faces were judged according to gender more slowly than adult faces, owing to their less differentiated facial characteristics, it is feasible that they cluster at the midpoint of the gender dimension with adult male and female faces pegging out the dimension.

Using the same stimuli as Experiment 1, plus jumbled versions of each face, Experiment 2 evaluated these predictions with a facedness task. A main effect of gender emerged stemming from the slower classification of male faces compared with female faces regardless of age. Given that classifications of males in Experiment 1 were not impeded by close exemplars, the slower facedness judgments of adult males were not surprising. Given the proximity of female adult faces and child faces to competing representations, the faster facedness decisions to female adult and female children was not surprising either. However, the longer facedness decision latencies to male child faces countered the prediction derived from Experiment 1 which considers them to overlap with female child faces. Johnston et al. (1997) enlisted an exemplar-based MDS to account for the slow classification of male child faces. They proposed that representations of child faces are particularly sparse thus slowing facedness decisions. However, because female child faces overlap with both male contemporaries and female adult faces, the disadvantage imparted by low density, is offset. Johnston et al. confirmed the hypothesis derived from Experiment 1: gender is coded by a dimension with androgyny at the midpoint and masculinity and femininity at either end.

O'Toole et al. (1998) employed factor analysis and PCA to explore two alternative mechanisms underlying gender classification in face-space. Like Johnston et al. (1997), the present study argued that face-space is characterised by two gender-based clusters. They reasoned that the space close to the average male and average female should be the area of highest exemplar density, therefore, recognition should depend on distance from the subcategory prototype[8]. If gender classification involves a comparison with the subcategory prototype, then recognition should be inversely related to classification latency (this prediction deems gender classifications analogous to facedness decisions). Alternatively, O'Toole et al. hypothesised that there could be a role for subcategory caricatures that express the maximally contrastive elements of categories (masculinity or femininity). Therefore, recognition and classification may dissociate with classification related to the masculinity or femininity of the face.

Factor analysis was carried out on recognition memory task and gender classification performance measures, and ratings of masculinity or femininity and attractiveness. PCA was subsequently implemented to assess how the relationship between ratings and performance measures related to stimulus structure differences between male and female faces. In both analyses, 'recognition' and 'classification' were extracted as the two major axes. O'Toole et al. (1998) reasoned that the independence of these axes negates the contention that that recognition and classification performance both depend on distance from the subcategory prototype. The analysis also revealed a differential loading of ratings onto each axis for male and female faces. Attractiveness and femininity loaded onto the classification axis for females. The size of the loadings was similar indicating that observers used both ratings in similar ways. O'Toole et al. noted that this sits uncomfortably with their conceptualisation of femininity as 'caricature' and the widely acknowledged conception of attractiveness as 'average' (e.g. Langlois & Roggman, 1990). The authors resolved 'caricature' as a better descriptor of

[8] O'Toole et al. (1998) do not explicitly adopt norm-based or exemplar-based versions of MDS. It seems that 'prototype' can be interchanged with 'area of highest exemplar density' at no cost to this study.

information carried by femininity and attractiveness ratings: they reasoned that because attractiveness did not load onto the recognition axis, it is unlikely to reflect 'averageness'. Attractiveness and masculinity also loaded onto the classification axis for male faces. However, the difference in the size of the loadings suggested that attractiveness was not synonymous with masculinity. Indeed, masculinity loaded onto the classification axis more strongly than attractiveness. The latter also loaded onto the recognition axis. O'Toole et al. deduced that a component of male attractiveness captures 'averageness' and masculinity captures 'caricature'.

O'Toole et al. (1998) concluded that recognition memory performance is concerned with the subcategory prototype since it is related to the number of similar competing exemplars. Conversely, classification is dependent on opposition to a contrastive category so a caricatured female face exaggerates features which distinguish it from male faces. On the other hand, Johnston et al. (1997) liken gender decisions to facedness decisions which are inversely related to recognition and bereft of any caricatured component. Both studies appear to be compatible with contemporary face-space models but modifications may be needed to implement O'Toole et al.'s (1998) notion of 'caricature'.

1.6 Outline of Forthcoming Chapters

Chapters 2 and 3 present experiments which seek to ascertain the nature of the relationship between AoA and distinctiveness; the core objective being to refine mechanisms underpinning AoA and shed light on associated loci. In the subsequent experiments, AoA is recruited as an investigative tool with which to assess the relationship between identity and gender processing. Chapter 4 presents an experiment which investigates the impact of AoA and distinctiveness on gender processing. Chapters 5 and 6 then offer a series of experiments which explore the effect of identity on gender processing and conditions which elicit this effect. In Chapter 7, experiments are presented which seek to constrain mechanisms proposed to subserve the relationship between identity and gender processing. A summary of the main findings and their implications for theories of AoA and face processing is subject to discussion in Chapter 8.

CHAPTER 2 – THE RELATIONSHIP BETWEEN AOA AND DISTINCTIVENESS

2.1 Database Creation and Collection of Ratings

A database containing current images of famous faces formed the stimulus base for Experiments 1, 2, 4, 5, 6, 9, and 10. Experiments 3, 7 and 8 were carried out with dated versions of these images, hence the additional creation of a dated image database.

Both databases comprised good quality images of famous faces, whether television presenters, television and film stars, pop singers, politicians, sports personalities or royalty. The current database comprised 260 up-to-date images and the dated version, 199 dated images of personalities featured in the current database. The dated images captured the faces as they looked when first acquired by the undergraduate cohort tested: early acquired faces were depicted as they appeared in the early 1980's and late 1990's. If the dated image depicted the personality as a child, for example Prince William, it was not included in the database, owing to the nonlinearity of changes that occur through ageing.

The images were obtained by scanning photographs from magazines, or obtained as digital stills from the internet or courtesy of PAPhotos picture library, see Figure 2.1. All images captured full-face depictions and were selected to ensure minimal variation in pose and expression. Using the Adobe PhotoShop 7.0 software package, the images were digitally edited to obscure background and clothing, and were cropped below the neck. The images were pasted onto a white background and equated in size to approximately 350 x 350 pixels, determined by approximate inter-ocular distance.

Figure 2.1. *Examples of edited stimuli.*

2.1.1 Current Database: Ratings

The database was rated for AoA, distinctiveness, frequency[9], likeness and familiarity by York University undergraduates, who had been brought up in the UK for at least 18 years. The collection of ratings spanned three separate sessions, employing different participants each time. Two hundred faces were rated for AoA, distinctiveness, frequency and likeness during the first session. However, after this session it was evident that ratings of AoA and distinctiveness were not evenly spread: a disproportionately small number of images were rated as early acquired and typical. Given that the first set of experiments investigate the relationship between these two variables, it was imperative to ensure a comprehensive range of stimuli. The database was extended in session two and a further 60 faces rated for AoA, distinctiveness, frequency and likeness. After the first experiment it became apparent that some of the famous faces featured may not have been sufficiently familiar to all participants. To combat this, faces were rated for familiarity in session 3 and screened for familiarity in subsequent experiments.

In session one, 21 participants (mean age = 19.38 years old, SD = .49) generated AoA and likeness ratings for 200 images, and 20 participants (mean age = 19.45 years old, SD = .51) generated distinctiveness and frequency ratings. The decision to get AoA and frequency rated separately stemmed from the controversial status

[9] Frequency of encounter was intended to be analogous to word frequency as there is no direct equivalent to word frequency in the face recognition literature. Moore and Valentine (1998) controlled for familiarity, but Lewis (1999b) pointed out that the subjectivity of this variable may mean that it is more likely to reflect recent occurrences than earlier ones.

of the 'familiarity' variable in face recognition literature. Familiarity can be interpreted as a proxy for cumulative frequency (Lewis, 1999a), which is itself a product of AoA and frequency. This inextricable link dictated the decision to rate the variables separately.

The ratings were carried out using PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993) for the AppleMac. The rating tasks were counter-balanced and item presentation randomized during each task. Participants were given as much time as required and permitted to break between the tasks. Throughout the tasks, the rating scale remained on the screen, and when rating AoA, frequency and likeness, the name of the face was presented with the image. Distinctiveness ratings were generated from the image alone to ensure that participants rated the distinctiveness of the face, rather than the person. Participants registered their rating with keys 1 to 5 or if a face was unknown, the space-bar. However, the option to select 'unknown' was not available when rating distinctiveness because these judgments can be made to familiar and unfamiliar faces and there is no evidence to suggest that they have a different origin (Valentine & Bruce, 1986a). Details of the rating instructions and scales are provided:

AoA. Participants were required to estimate the age at which they first became aware of a celebrity, using a 5-point scale. As an aid, participants were told to contemplate "whether you were at primary school or secondary school at the time". As a further means of facilitating performance, cues were written alongside each point on the scale to contextualise childhood (1 = less than 5 years, preschool; 2 = 5 – 8 years, early school; 3 = 9 – 12 years, middle school; 4 = 13 – 16 years, secondary school; 5 = over 16 years, post GCSEs).

Distinctiveness. Participants were asked to rate faces according to how typical or distinctive they appeared, using a 5-point scale (1 = very typical, 2 = fairly typical, 3 = mildly distinctive, 4 = fairly distinctive, 5 = highly distinctive).

Examples of distinctive and typical faces were provided with the instructions in order to facilitate an awareness of a range of face types[10].

Frequency. Ratings were made on a 5-point scale pertaining to how often the celebrity was encountered at the time of testing (1 = once a year or less, 2 = a few times a year, 3 = about once a month, 4 = about once a week, 5 = almost once a day).

Likeness. A rating of how well the image resembled the person concerned permitted a quality control of the images themselves. Ratings were made on a 5-point scale (1 = very poor, 2 = poor, 3 = reasonable, 4 = good, 5 = very good).

In session two, 23 participants (mean age = 18.61 years old, SD = .72) generated AoA and likeness ratings for 60 images, and 23 participants (mean age = 18.48 years old, SD = .51) generated distinctiveness and frequency ratings. Instructions and rating scales were the same as those employed previously. However, on this occasion the ratings tasks were administered as a booklet with a cover page detailing instructions and a rating scale printed on the top of each page. Participants were informed to complete the tasks in the order specified. The 120 images were spread over 6 pages and items were pseudo-randomized between participants in that the pages were ordered differently.

In session 3, the entire database (n = 260) was rated for familiarity to allow the removal of faces which were not at the very least deemed 'familiar'. Twenty two participants (average age = 18.27, S.D. = .55) were asked to rate each face with respect to how familiar it was to them (1 = unfamiliar; 2 = vaguely familiar; 3 = fairly familiar; 4 = familiar; 5 = very familiar). The task was administered in booklet format with a cover page detailing instructions and the rating scale printed on the top of each page. The 260 items were spread over 11 pages and participants were asked to complete the booklet in the order directed. Items were pseudo-randomized between participants in that the pages were ordered differently.

---

[10] Hosie and Milne (1995) articulated the notions of 'absolute distinctiveness', which emerges with experience, and 'relative distinctiveness', referring to effects which arise as a function of the experimental stimulus set. Providing illustrated examples is one way of minimising relative distinctiveness effects.

The split-half reliability of each scale was checked by correlating the mean ratings of half of the participants with the mean ratings of the remaining half. It is interesting to note that the ratings of the non-visual properties of the stimulus, AoA ($r = .91$), frequency ($r = .90$) and familiarity ($r = .91$), were highly reliable. Conversely, ratings made to visual attributes, distinctiveness ($r = .75$) and likeness ($r = .60$) were less reliable. The particularly poor likeness coefficient is discussed shortly.

Correlations between AoA, distinctiveness, frequency, likeness and familiarity are shown in Table 2.1. The analyses reported were performed using two-tailed Pearson's Correlation Coefficients. Six significant correlations emerged from the analysis. Of the moderate correlations, faces rated as more frequently encountered tended to be rated as later acquired, $r(258) = .43$, $p<.01$, and more familiar, $r(258) = .59$, $p<.01$. Also, images of familiar faces were deemed better depictions, or likenesses, compared with their less familiar counterparts, $r(258) = .49$, $p<.01$. This is surprising as there are no reasons to suspect that intrinsic familiarity with a face should be related to image quality. It is quite possible that when participants felt less familiar with a face, they attributed this feeling to the 'poor' image, or perhaps poor quality depictions attenuated the feeling of familiarity. Either way, if at least some participants are associating likeness with familiarity, particularly at the 'unfamiliar' end of the scale, this may explain why the likeness reliability coefficient is low. The split-half reliability of the likeness scale was assessed once again, but this time faces low in familiarity were removed. Given that the database was screened to ensure that all faces were at least 'familiar' or 'very familiar' after the first experiment, it seemed sensible to gauge the reliability of likeness ratings for the database ($n = 213$) used after the first experiment. Indeed, the likeness reliability coefficient improved ($r = .78$) once faces low in familiarity were discarded.

Table 2.1. *Correlation Matrix for the 5 Variables in the Analysis of the Current Database (n = 260).*

	AoA	Dist.	Frequency	Likeness	Familiarity
AoA	1.00				
Dist.	-0.10	1.00			
Frequency	0.43**	0.01	1.00		
Likeness	0.05	0.17**	0.36**	1.00	
Familiarity	0.05	0.25**	0.59**	0.49**	1.00

**. Correlation is significant at the 0.01 level (two-tailed)

Of the correlations considered to be weak, faces rated as familiar were also rated as more distinctive, $r(258) = .25$, $p<.01$, echoing Moore and Valentine's (1998) correlation between the two variables, $r(99) = .27$, $p<.01$. Also, faces rated as more frequently encountered were considered to be better depictions of the face, $r(258) = .36$, $p<.01$. There were no reasons to expect this correlation other than those suggested to underpin a similar relationship between likeness and familiarity. Likeness was also very weakly correlated with distinctiveness, $r(258) = .17$, $p<.01$. Despite the fact that quality of resemblance should be independent of structural facial characteristics, the finding that better likeness ratings are awarded to distinctive faces is analogous to the caricature advantage (e.g. Chang, Levine & Benson, 2002) insofar as distinctive features yield better likeness judgments.

Moore and Valentine (1998) found that faces rated as early acquired were rated as more familiar, $r(99) = -.26$, $p<.01$ and more distinctive, $r(99) = -.37$, $p<.01$. These relationships did not emerge from the present analysis. Whereas their database was created for entry into a multiple regression analysis, efforts were made to ensure an equal spread of AoA and distinctiveness ratings to facilitate a factorial analysis of these variables. It is possible that the lack of correlation between AoA and distinctiveness may be a product of the criteria fulfilled to enable this analysis.

2.1.2 Dated Database: Ratings

The dated database was rated for distinctiveness and likeness by York University undergraduates who had not provided ratings for the current database. AoA and frequency were not rated again because non-visual properties of the stimulus should not vary with facial age. Twenty seven participants (mean age = 18.67, S.D. = .68) generated distinctiveness ratings for the 199 images and 26 of the same participants (mean age = 18.65, S.D. = .69) generated likeness ratings. The task was administered as a booklet with a cover page detailing instructions and the rating scale printed on the top of each page. The 199 items were spread over 8 pages and participants were asked to complete the booklet in the order directed. Items were pseudo-randomized between participants in that pages were ordered differently. Details of distinctiveness rating instructions and scales did not differ from those provided when rating current images. However, likeness rating instructions were changed to accommodate differences in the images. Participants were informed that they would see photographs of famous faces as they looked a number of years ago. They were instructed to decide if the photograph was a good resemblance of the person as they would have looked when the photograph was taken. Ratings were made using a 5-point rating scale (1 = very poor likeness; 2 = poor likeness; 3 = reasonable likeness; 4 = good likeness; 5 = very good likeness).

The split-half reliability of the scales was checked. The distinctiveness ratings ($r = .87$) and likeness ratings ($r = .80$) were reliable. Moreover, the analysis yielded better coefficients than those associated with the current database. This may have arisen because the tasks were perceived as harder when applied to dated faces compared with current images, thus demanding more attention. The internal consistency of the likeness ratings indicated that participants could complete the task sufficiently well despite the demands involved in comparing the likeness of a dated percept with a dated canonical representation in memory.

Correlations between distinctiveness, likeness and the previously rated variables, AoA, frequency and familiarity are shown in Table 2.2. The analyses reported were performed using two-tailed Pearson's Correlation Coefficients. Five significant correlations emerged from the analysis. Not surprisingly, correlations between variables which had not been re-rated remained significant: faces rated as more frequently encountered were rated as later acquired, $r(197) = .24$, $p<.01$, and more familiar, $r(197) = .62$, $p<.01$. Of the new ratings, two weak correlations emerged. Faces rated as distinctive were granted better likeness ratings than typical faces, r(197) = .24, p<.01, and rated as more familiar, r(197) = .21, p<.01, preserving the relationships seen in the current database. Finally, a very weak relationship between likeness and familiarity emerged, again mirroring that found with current images, r(197) = .17, p<.05. Whereas frequency of encounter was related to the likeness attributed to current images, this correlation was not significant when applied to dated images. Demonstrated by the improved reliability coefficient, it is possible that the likeness ratings to dated faces may be tapping measures of resemblance more effectively than those made to current images. This could explain the lack of correlation between frequency of encounter and likeness, and is also consistent with the very weak correlation between likeness and familiarity which was moderately correlated beforehand.

Table 2.2. *Correlation Matrix for the 5 Variables in the Analysis of the Dated Database (n = 199).*

	AoA	Dist.	Frequency	Likeness	Familiarity
AoA	1.00				
Dist.	-0.14	1.00			
Frequency	0.24**	-0.09	1.00		
Likeness	-0.12	0.24**	-0.12	1.00	
Familiarity	-0.08	0.21**	0.62**	0.17*	1.00

**. Correlation is significant at the 0.01 level (two-tailed)
*. Correlation is significant at the 0.05 level (two-tailed)

2.2 Introduction: Experiments 1 to 3

The variables AoA and distinctiveness exert robust effects in the domain of face recognition and yet they have never been investigated in conjunction with one another. This is not surprising because intuition would suggest that the variables are independent. In Lewis' (1999a) terms, whereas AoA is a non-pictorial variable, distinctiveness is pictorial and refers to properties of the visual stimulus. In addition, whilst AoA effects are clearly temporal in origin, distinctiveness effects arise as a result of the spatial distribution of representations in memory. Both functionally and mechanistically, there are not obvious reasons to suppose interdependence. However, a rationale for linking these variables stemmed from the notion that distinctiveness effects emerge as a function of experience. It is possible that the status of distinctiveness effects at a particular point in time are duly imprinted in the representations of faces acquired at that time. Experiments 1 to 3 sought to ascertain the nature of the relationship between AoA and distinctiveness with the aim of shedding light on the mechanisms underpinning both variables. This section certifies that both variables are capable of supporting an interaction; a necessary prerequisite to defying intuition. A theory which integrates the mechanisms underpinning AoA and distinctiveness is then outlined.

It is feasible that mechanisms yielding AoA effects could support this interaction. AoA effects are not idiosyncratic to one particular locus and as such, underlying mechanisms are unlikely to reflect the specificity associated with certain loci. Moore and Valentine (1999) hypothesised that the variable could be a general property of the representation of perceptual and lexical information. Indeed, studies have shown that AoA affects familiarity decisions (Moore & Valentine, 1999), semantic decisions (Lewis, 1999a) and face naming (Moore & Valentine, 1998), reflecting operations at key face processing stages. Further, A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis proposes that AoA effects are a fundamental by-product of cumulative and interleaved learning. Hence, there are no *a priori* reasons to suspect this interaction untenable.

The forthcoming hypothesis is couched in terms of A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis and the learning algorithm underpinning this implementation is pivotal in permitting a theoretical integration of AoA and distinctiveness. Therefore, a recapitulation of this algorithm is provided to facilitate understanding of the hypothesis. The connectionist network is trained by a back-propagation algorithm. This entails the computation of an error signal which corresponds to the difference between the pattern of activation at the output units and the desired target pattern. This signal is propagated back through the network causing iterative adjustments in the weights connecting the units. At the beginning of training, unit activations are set to 0.5, reaching 0 or 1 during training. As performance reaches asymptote, the rate of weight change reduces causing a reduction in network plasticity. Consequently, early patterns impart greater influence on the structure of the mappings between input and output units than patterns trained later. This reduction in plasticity provides the mechanistic link with distinctiveness.

The temporal dimension of distinctiveness allows integration with this reduction in network plasticity. Valentine (1991) hypothesised that distinctiveness effects emerge as a result of the centrally-clustered distribution of representations in multidimensional face-space, and that this distribution evolves as a function of experience of facial encounters. This is consistent with H. Ellis (1992) and Johnston and H. Ellis' (1995) finding that distinctiveness effects do not begin to emerge in recognition memory tasks until around 9 years of age. Various conceptualisations of MDS were proposed to chart its evolution from child spaces to the centrally-clustered adult space, thus demonstrating that distinctiveness does have a temporal dimension. It is possible that the evolving state of MDS over time integrates with the loss in plasticity posited to underpin AoA effects.

The hypothesis offered presently suggests that the differential effect of distinctiveness over time is manifested in the representations of early and late acquired faces. Early acquired faces (analogous to network patterns trained early) have the opportunity to structure MDS (analogous to the network) towards their preferred configuration. MDS becomes progressively attuned to early acquired faces and less receptive to faces acquired later. This reduction in plasticity not only configures MDS favourably towards early acquired faces, but also imprints the distribution of these faces at the time of acquisition. Consequently, the attenuated distinctiveness effect associated with child face-space[11] is maintained by the representations of early acquired faces.

This hypothesis demands an elaboration of mechanisms underpinning A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis and Valentine's (1991) MDS in order to explain an interaction between AoA and distinctiveness. If the absence of distinctiveness effects in childhood and the presence in adulthood is manifested

[11] Although H. Ellis (1992) proposed uniform and differential versions of child face-space, the two versions are interchangeable as far as the present hypothesis is concerned: this hypothesis could be grafted onto either conceptualisation. In a recent review, Valentine (2001) favoured exemplar-based uniform versions of MDS, therefore the present hypothesis employs the uniform child face-space to explain hypothesised mechanisms.

in the respective representations of early and late acquired faces, then it is hypothesised that Valentine's MDS comprises discrete layers of space housing representations encoded at different points in time (see Figure 2.2). Additionally, it is logical to suppose that AoA effects are confined to the encoding and storage of typical faces and therefore, it is hypothesised that loss of network plasticity affects distinctive and typical faces differently. As an increasing number of faces are assimilated into MDS over time, competition for space in central 'typical' regions is exacerbated and recognition of late acquired typical faces is compromised relative to their early acquired counterparts. Conversely, the recognition of distinctive faces may be less susceptible to AoA effects because less competition for space in peripheral regions ensures that loss of plasticity imparts less of an impact. This is analogous to previous studies which have demonstrated that the network structure formed by early patterns is hostile to the assimilation of later acquired patterns under certain circumstances and other circumstances in which it is conducive. The interaction between AoA and word consistency is one such example (A. Ellis & Lambon Ralph, 2000; Monaghan & A. Ellis, 2002; Zevin & Seidenberg, 2002).

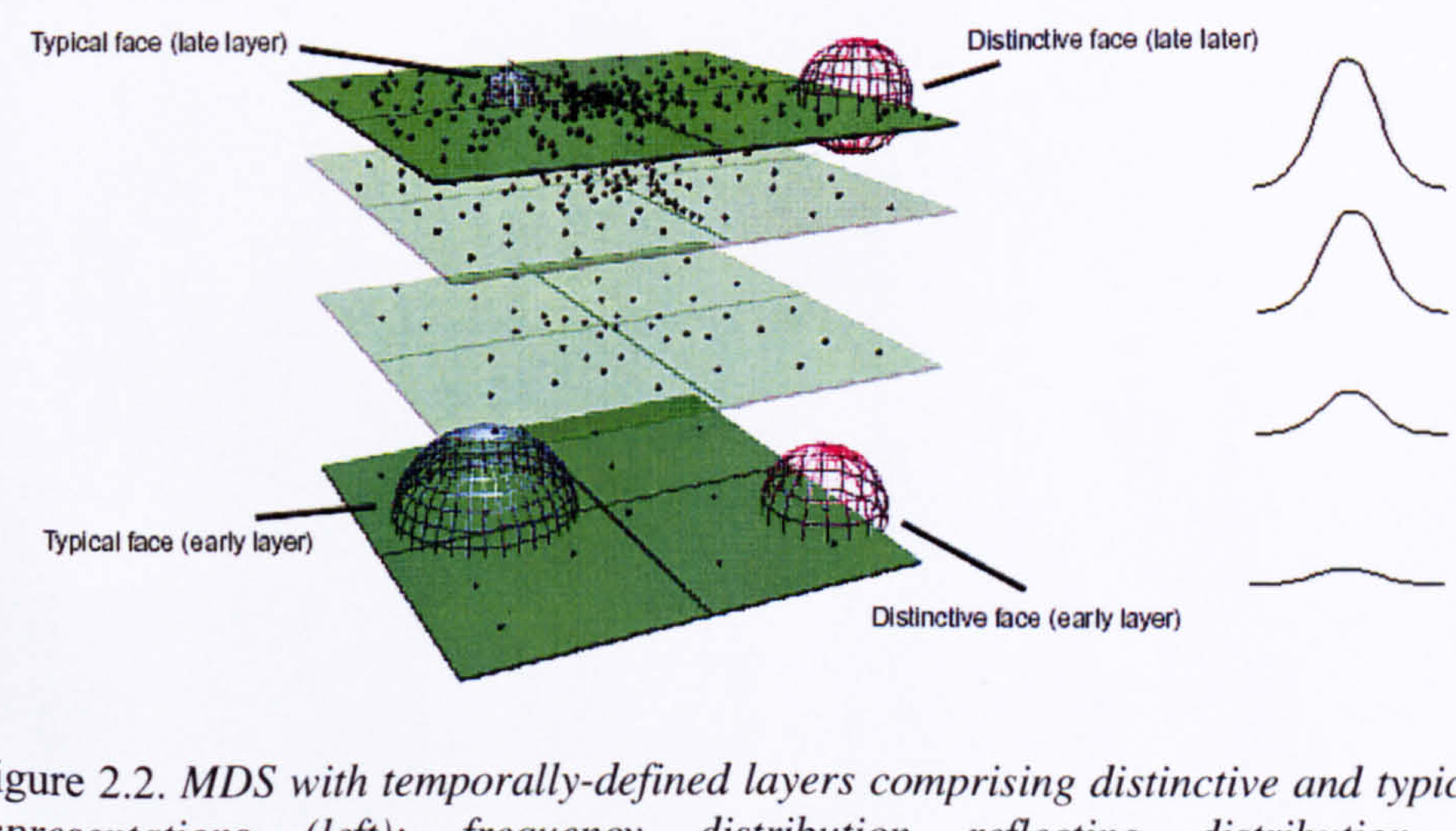


Figure 2.2. *MDS with temporally-defined layers comprising distinctive and typical representations (left); frequency distribution reflecting distribution of representations in MDS at different points in time (right).*

It is possible that intuition will prevail and an additive relationship between AoA and distinctiveness will emerge, signalling independence between the variables. Although disparaged these days and largely undermined by cascade models, additive factors logic (Roberts & Sternberg, 1993; Sternberg, 1969; 1998) is tentatively recruited for the light that it can shed upon the locus of action of these variables, should AoA and distinctiveness be shown to be additive. Using the same rationale as Bruce, Burton and Dench (1993b) who interpreted additive effects of distinctiveness, repetition priming and semantic priming to indicate that these variables operate at serially-arranged processing stages, additivity between AoA and distinctiveness would suggest that the variables act at different processing stages. Given that distinctiveness effects are linked to FRU activity (Burton et al., 1990; Newell et al., 1999) this would suggest that AoA does not affect mechanisms at this early stage in face recognition. Furthermore, Bruce et al. (1993b) noted that the pattern of independent effects "is difficult to accommodate within a model of face processing in which familiarity decisions are taken at the level of face recognition units" (p. 57). Indeed, the most parsimonious account of the data is a PIN-level familiarity decision mechanism. Similarly, if distinctiveness affects FRU activity and AoA affects subsequent processing loci, it is most parsimonious to conclude that familiarity decisions are taken at the level of PINs. This would provide further support for the contention that PINs house familiarity decision mechanisms.

Relatively little is known about the impact of age-related change on the encoding and storage of faces for recognition. O'Toole et al. (1997) discovered that caricaturing 3D laser scanned heads increased their perceived age. Moreover, Deffenbacher et al. (1998) explicitly claimed that faces become increasingly distinctive with age. Thus, it is conceivable that representations move away from the origin of MDS with age. For the hypothesised interaction between AoA and distinctiveness to emerge, the state of MDS at the point of acquisition must be manifested accordingly in the representations of early and late acquired faces. It is possible that the interaction might be enhanced if images captured faces when first encountered and encoded into MDS. Whilst Experiments 1 and 2 pursued this investigation with up-to-date images of famous faces, Experiment 3 used images which captured the faces as they appeared when first acquired.

An interaction between AoA and distinctiveness was hypothesised, reflecting an integration of mechanisms. Distinctiveness was predicted to affect late acquired faces but not early acquired faces, echoing the maturational attenuation of distinctiveness effects at the time of acquisition. AoA was predicted to affect typical faces but not distinctive faces, stemming from the differential impact of loss of network plasticity. A pilot experiment tested for the plausibility of this interaction using an existing database of famous face images before embarking on Experiments 1 to 3.

2.3 Pilot Experiment

2.3.1 Introduction

The pilot experiment tested for the plausibility of an interaction between AoA and distinctiveness, taking advantage of an existing database of famous face images. This ensured that no time would be wasted creating a new database specifically tailored to meet task demands, when the search for an interaction may prove fruitless.

Familiarity decisions were made to sets of faces that were either early acquired and distinctive, early acquired and typical, late acquired and distinctive or late acquired and typical. The familiarity decision task, developed as an analogue of the lexical decision task (Bruce, 1983), was chosen to investigate the influence of AoA and distinctiveness on recognition. This task by-passes the "more effortful and error prone process of name retrieval" (Burton et al., 1990, p.362) and activates stored representations without interference from semantic activation. Moreover, there are no *a priori* reasons to assume that AoA and distinctiveness operate at selective stages of processing or at a higher-order stage, respectively.

In line with robust AoA and distinctiveness effects documented in the literature, superior processing of early acquired faces compared with late acquired faces was predicted, as was the recognition advantage for distinctive faces compared with their typical counterparts. In particular, an interaction between AoA and distinctiveness was hypothesised, with distinctiveness effects restricted to late acquired faces and AoA effects restricted to typical faces.

2.3.2 Method

Participants. Nineteen York University undergraduates (17 female, 2 male) took part in the experiment, receiving course credit or a payment of £2 for their participation. Recruitment was contingent upon the satisfaction of two criteria: 1) participants had to be between 18-25 years old; this constituted the range tested; 2) participants had to have been exposed to the British media for at least the first 18 years of their life. The mean age of participants was 20.95 years old (SD = 1.81), approximately the same age as participants who rated the stimuli prior to this experiment

Design. The experimental variables comprised AoA (with two levels, early and late) and distinctiveness (with two levels, typical and distinctive). The variables were manipulated factorially in a 2 x 2 within-subjects design yielding the following four sets: early acquired distinctive; early acquired typical; late acquired distinctive; late acquired typical. Response latency and accuracy were the products of face familiarity decision task.

Materials. Images of famous faces were extracted from a pre-existing database (Rayner & A. Ellis, personal communication). The database had been created by scanning 300 full-face images of famous faces and adjusting them to approximately 350 x 350 pixels in size. These images had been digitally edited to obscure background and clothing, and were cropped below the neck. The images had been previously rated for AoA, distinctiveness, frequency and likeness by different groups of 20 undergraduate students, who had been brought up in the UK. Images were projected onto a screen one at time and participants noted their responses on answer sheets. Ratings were made on a 5-point scale, and with exception to distinctiveness, participants had the option to select 'unknown'. Details of the distinctiveness and likeness rating instructions and scales did not differ from those used in the creation of the current database (see p. 75). Whilst the

AoA and frequency instructions did not differ, the AoA scale varied slightly (1 = less than 6 years; 2 = 7 – 10 years; 3 = 11 – 14 years; 4 = 15 – 18 years; 5 = over 18 years) as did the frequency scale (1 = less than once per year; 2 = at least once per year; 3 = at least once per month; 4 = at least once per week; 5 = at least once per day).

To provide a measure of semantic knowledge of each celebrity, 20 participants were given one minute to write down as many facts as they could per celebrity. Facts were accepted regardless of accuracy as all were reasoned to reflect the existence of individual semantic units.

Sixty four faces (16 early acquired distinctive, 16 early acquired typical, 16 late acquired distinctive, 16 late acquired typical) were selected from the database. The sets were determined by splitting the AoA and distinctiveness ratings at the median and extracting 16 faces from each sector. Independent t-test confirmed that the sets were matched for frequency, likeness and average number of facts, and that the two distinctive and two typical sets were each matched for distinctiveness and that the two early sets were matched for AoA[12] (in each case $t \leq 1.62$). See Table 2.3 for summary data pertaining to the creation of experimental sets.

To facilitate a familiarity decision task, famous faces were matched with an equal number of unfamiliar faces by approximate age, gender and race. The unfamiliar faces were digitally treated in the same manner as the famous faces.

Procedure and Apparatus. The familiarity decision was run on PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to the images which occupied the centre of the screen. Each experimental trial began with a 500 ms delay prior to image onset. Faces were displayed until a response was executed. 'Familiar' and 'unfamiliar' responses were logged using the 'z' and '/' keys on the keyboard, and

[12] The late acquired distinctive and late acquired typical sets were not statistically matched for AoA, $t(30) = -2.20$, $p<.05$. However, a covariate analysis of variance carried out on the forthcoming late acquired dataset demonstrated that unwanted variation in AoA did not adversely affect the latency, $F_2(1,29) = 0.30$, $MSe = 7,662.03$, $p = .59$, or accuracy data, $F_2(1,29) = 0.90$, $MSe = 3.78$, $p = .76$.

keys pressed with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 pixels and their absolute dimensions subtended 70 x 150 mm approximately. Participants viewed these from a distance of approximately 30cm.

The 128 faces were presented in a single block, owing to the short duration of the experiment. Presentation order was randomized between participants. Twenty practice trials (10 famous and 10 unfamiliar) were presented before the experimental trials, after which participants were reminded to respond as quickly and accurately as possible. The experimental trials were directly preceded by four lead-in trials as a means of concealing initial hesitations. Both practice and buffer trials contained a representative sample of stimuli (early and late acquired, typical and distinctive, famous and unfamiliar, male and female) to preclude potential bias effects. None of these items were repeated as experimental items, neither were they included in the analysis. The task lasted approximately 10 minutes.

Table 2.3. *Mean Ratings as a Function of Experimental Set (Pilot Experiment).*

	Early acq. & distinctive			Late acq. & distinctive			Early acq. & typical			Late acq. & typical		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.41	(0.46)	1.50 - 2.88	3.33	(0.38)	2.94 - 4.18	2.45	(0.39)	1.32 - 2.90	3.64	(0.42)	3.06 - 4.52
Distinctiveness	3.97	(0.50)	3.30 - 4.90	3.72	(0.34)	3.25 - 4.30	2.89	(0.32)	2.25 - 3.25	2.77	(0.32)	2.10 - 3.25
Frequency	2.66	(0.65)	1.75 - 3.80	2.84	(0.68)	2.05 - 4.42	2.8	(0.57)	1.94 - 4.20	2.77	(0.60)	1.95 - 3.75
Likeness	4.05	(0.47)	3.00 - 4.70	4.07	(0.40)	3.35 - 4.90	4.02	(0.72)	2.63 - 4.85	3.94	(0.52)	3.10 - 4.60
Av. No. Facts	6.61	(1.17)	4.60 - 8.40	6.53	(1.20)	4.85 - 8.25	6.62	(0.98)	5.10 - 9.00	6.64	(1.32)	4.85 - 9.20

2.3.3 *Results*

Prior to the analysis, three participants were replaced because either they failed to recognise at least 75% of the famous faces, or 75% of the famous and unfamiliar faces when pooled together. The first criterion was implemented to ensure that the latency analysis was based on a sufficient number of responses and the second, to prevent inclusion of data derived from adverse response strategies.

The accuracy analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions; b) reaction times deemed outliers. As a result of incorrect familiarity decisions made to famous faces, 65 (5.4%) of the responses were removed. As a result of eliminating outliers slower than 3 standard deviations away from the mean and faster than 300ms, 13 (1.1%) of the responses were discarded. However, the accuracy analysis should be addressed with caution because it is impossible to distinguish between a genuine error and unfamiliarity with a target face, thus limiting its validity.

The latency analysis was carried out on correct responses only. Table 2.4 shows the mean correct response latencies in each condition of the experiment and percentage error rates. Unfamiliar face summary data is also reported. For inspection purposes, '% misclassifications', referring to the percentage of incorrect familiarity decisions, are documented.

Table 2.4. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Pilot Experiment).*

	Early Acquired		Late Acquired		Unfamiliar
	Distinctive	Typical	Distinctive	Typical	
Mean RT	847	854	871	966	1014
SD	(148)	(156)	(167)	(176)	(257)
% Error	4.93	5.92	3.95	10.86	4.93
% Misclass.	4.28	4.61	2.96	9.54	2.55

Reaction Times. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors[13]. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of AoA was significant in both analyses, $F_1(1,18) = 23.46$, MSe = 3,750.61, $p<.001$; $F_2(1,60) = 6.21$, MSe = 10,723.22, $p<.05$, stemming from the faster recognition of early acquired faces (851 ms) compared with late acquired faces (919 ms). The main effect of distinctiveness was also significant in both analyses, $F_1(1,18) = 6.21$, MSe = 7,977.94, $p<.05$; $F_2(1,60) = 4.56$, MSe = 10,723.22, $p<.05$, adhering to the customary advantage found for distinctive faces (859 ms) compared with typical faces (910 ms).

The interaction between AoA and distinctiveness was significant in the by-subject analysis alone, $F1(1,18) = 5.77$, MSe = 6,312.79, $p<.05$; $F2(1,60) = 2.25$, MSe = 10,723.22, $p = .14$, see Figure 2.2. With alpha set at .05, *a posteriori* Tukey comparisons (HSD[14] = 73) demonstrated that typical faces were affected by AoA, whereas distinctive faces were unaffected. Further, distinctiveness affected the recognition of late acquired faces, but not those acquired early.

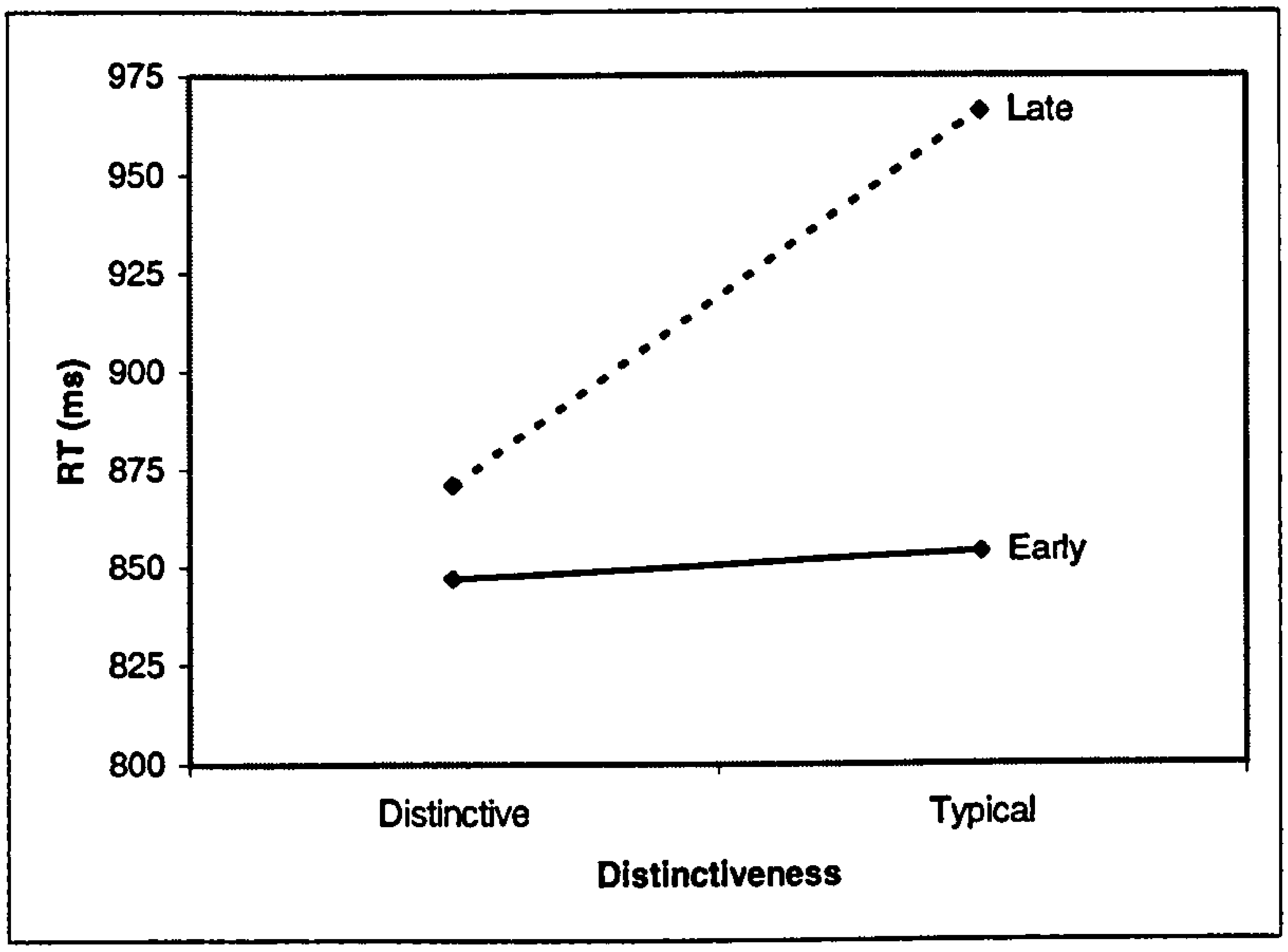
[13] By-subject analyses are given more weight than by-items analyses in data interpretation as there are *a priori* reasons to expect attenuated effects in by-items analyses: this sets the precedence for the entire thesis. Justification stems from the following: 1) the notion that there is no need to perform separate by-subject and by-item analyses, when item variability is experimentally controlled by matching variables that strongly correlate with the dependent variable, has encountered reappraisal (Raaijmakers, Schrijnemakers & Gremmen, 1999; Raaijmakers, 2003). The authors state that if items are matched between conditions, the variability in the mean item effect between conditions is reduced compared with the variability of the item effect within conditions, rendering the item analysis too conservative; 2) commenting on situations where stimuli are assigned to groups on the basis of a median split, Moore and Valentine (1998) suggest that this design is only appropriate when mean latencies, pooled across items, are analysed by-subject. If latencies are calculated per item "there will be many items of early AoA that are acquired only slightly earlier than many of the late acquired items" and "one would not expect an effect of AoA in an items analysis" (p. 502).

[14] Tukey's Honestly Significant Difference (HSD) is derived from the following formula where *Q* is a Studentised Range Statistic (determined by the df for the error term and the number of means subject to comparison), *MSerror* refers to mean squared error, and *n* refers to the number of observations making up each mean.

$$HSD = Q\sqrt{\frac{MSerror}{n}}$$

The HSD value represents the threshold for deciding whether or not two means are significantly different. Differences greater or equal to HSD are considered to be significantly different.

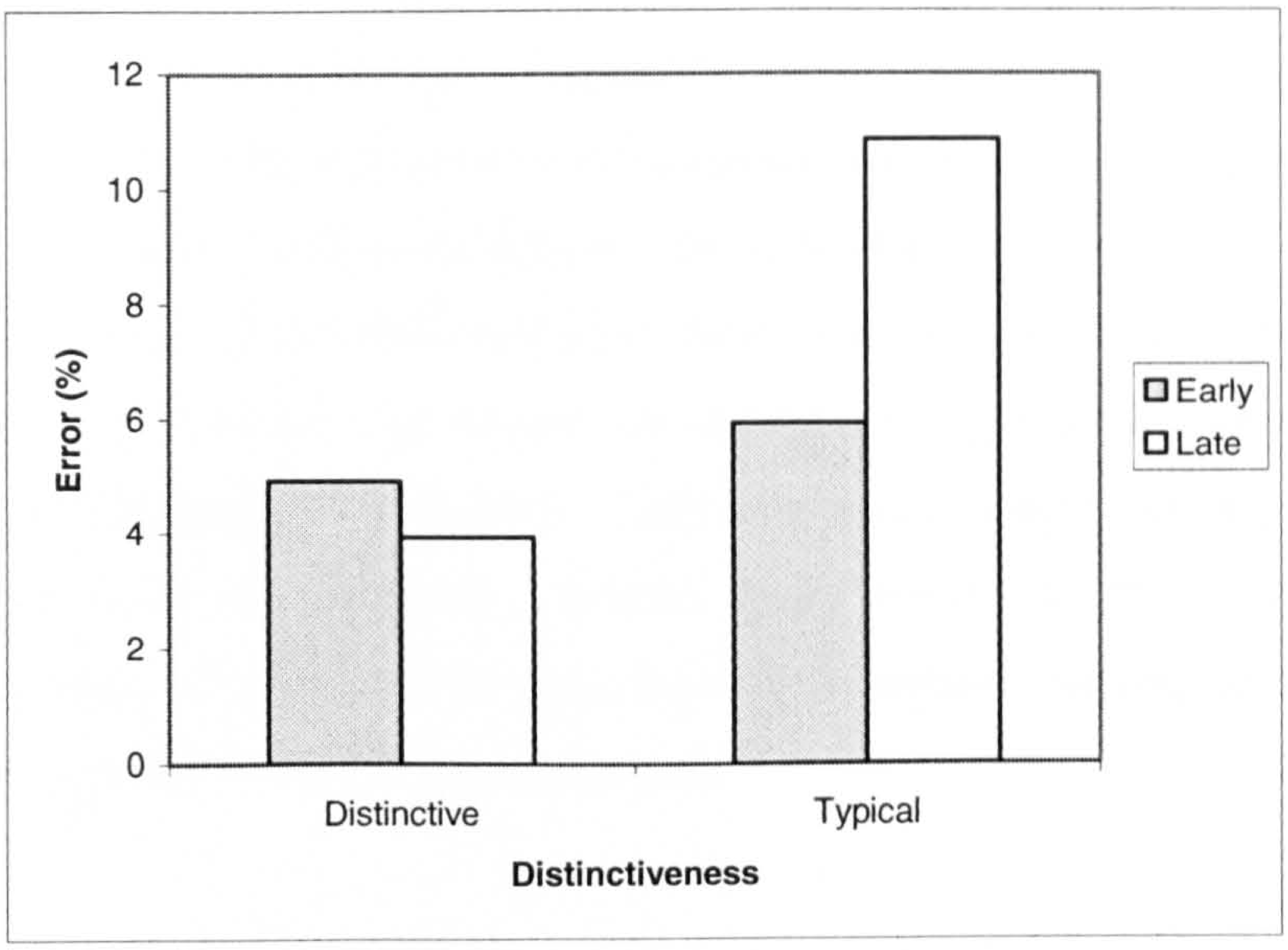
Figure 2.3. *Effect of AoA on Familiarity Decision RTs as a Function of Distinctiveness (Pilot Experiment).*



Errors. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. AoA was not significant in either analysis, $F_1(1,18) = 1.33$, MSe $= 1.42$, $p = 0.26$; $F_2(1,60) = 0.65$, MSe $= 4.05$, $p = .42$. The main effect of distinctiveness reached significance in the by-subjects analysis, $F_1(1,18) = 8.57$, MSe $= 0.89$, $p<.01$, confirming the tendency for distinctive faces (mean n = 0.7, 4.4%) to be recognised more accurately than typical faces (mean n = 1.3, 8.4%). The effect failed to attain significance in the by-items analysis, $F_2(1,60) = 2.81$, MSe $= 4.05$, $p = .10$.

A two-way interaction between AoA and distinctiveness was significant by-subjects only, $F_1(1,18) = 9.32$, MSe $= 0.46$, $p<.01$; $F_2(1,60) = 1.12$, MSe $= 4.05$, $p = .30$, see Figure 2.4. Tukey tests revealed that AoA affected typical face recognition accuracy, leaving distinctive faces unaffected, and late acquired faces were affected by distinctiveness, whereas early acquired faces were not affected (HSD = 0.62).

Figure 2.4. *Effect of AoA on Familiarity Decision Accuracy as a Function of Distinctiveness (Pilot Experiment).*



2.3.4 Discussion

The pilot experiment recovered robust effects of AoA and distinctiveness on the time taken to make familiarity decisions to famous faces. These effects were qualified by a significant interaction which assumed the form predicted: AoA affected the recognition of typical faces and the distinctiveness effect exclusively modulated performance associated with late acquired faces. The accuracy analysis yielded an effect of distinctiveness mirroring the latency analysis, but no AoA effect emerged. This is not particularly concerning given the afore mentioned caveat stating that it is impossible to distinguish between a genuine error and unfamiliarity with a target face, thus limiting the validity of the accuracy analysis. The accuracy analysis did, however, yield a significant interaction between AoA and distinctiveness mirroring that in the latency analysis and indicating that speed-accuracy trade-offs were not in operation. A conspicuous feature of the accuracy data was the reversed AoA effect found for distinctive faces. Although this finding was not statistically significant, it flagged potential issues with this database that needed to be taken into account when creating the new database.

Not all of the images in this database were current depictions of the faces. As a result, it is possible that AoA was confounded with age of image because images of late acquired faces are more likely to be recent than images of early acquired faces. Whilst it is accepted that faces are recognised despite age transformations (George & Hole, 1998), the effects of these transformations are not clear. It makes intuitive sense not to introduce this confound into the experiment. It is possible that more errors were made to the early acquired distinctive faces than the late acquired distinctive faces because the former set contained a greater number of dated images, attracting greater uncertainty. Taking this interpretation one step further, if early acquired distinctive faces attracted more errors for this reason, it is conceivable that reaction times to these faces were inflated, causing an artefactual interaction between AoA and distinctiveness.

CHAPTER 3 – THE RELATIONSHIP BETWEEN AOA AND DISTINCTIVENESS - CONTINUED

3.1 Experiment 1

3.1.1 Introduction

Promising results from the Pilot Experiment justified further exploration into the interaction between AoA and distinctiveness with a new database of current images, created specifically for this purpose. All images were up-to-date depictions of famous faces, thus standardising age of image. Again, familiarity decisions were made to sets of faces either, early acquired and distinctive, early acquired and typical, late acquired and distinctive or late acquired and typical. Again, the customary effects of AoA and distinctiveness were anticipated and the hypothesised interaction predicted.

3.1.2 Method

Participants. Twenty four undergraduate students from the University of York (19 female, 5 male) participated in this experiment, in return for £2 or course credit. Participation was contingent upon meeting the following criteria: a) participants had to be between 18 – 21 years old; b) participants were required to have lived in the UK for at least 18 years. The mean age of participants was 19.08 years old (SD = 0.58), and ages ranged from 18-20 years old. These age characteristics matched those of the participants who rated the database.

Design. This experiment took the form of a 2 x 2 within-subjects design, with the variables, AoA (with two levels, early and late) and distinctiveness (with two levels, distinctive and typical). These were manipulated factorially to yield early acquired distinctive, early acquired typical, late acquired distinctive and late acquired typical late conditions. A face familiarity decision task yielded two dependent variables; response latency and response accuracy.

Materials. A subset of 136 images from the current database formed the stimulus set. At this point in proceedings, images had been rated for AoA, distinctiveness, frequency and likeness. The four sets (34 early acquired distinctive, 34 early acquired typical, 34 late acquired distinctive, 34 late acquired typical) were created by splitting the face database in accordance with the median point of the AoA and distinctiveness ratings and extracting 34 faces from each sector. Independent t-tests confirmed that the sets were matched for frequency and likeness and that the two distinctive and two typical sets were each matched for distinctiveness and that the two early sets and two late sets were each matched for AoA (in each case $t \leq 1.48$). See Table 3.1 for summary data pertaining to the creation of experimental sets.

Famous faces were matched with an equal number of unfamiliar faces using gender, race and approximate age as controls. These unfamiliar faces were digitally treated in the same manner as the famous faces.

Procedure and Apparatus. The familiarity decision task was executed in PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to the images, which were presented in the centre of the screen. Each experimental trial began with a 500 ms delay prior to image onset. Faces were displayed until a response was made. Participants logged 'familiar' and 'unfamiliar' responses by pressing the 'z' and '/' keys on the keyboard with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 pixels and their dimensions subtended 70 x 150 mm approximately. Images were viewed from a distance of approximately 30cm.

Table 3.1. *Mean Ratings as a Function of Experimental Set (Experiment 1).*

	Early acq. & distinctive			Late acq. & distinctive			Early acq. & typical			Late acq. & typical		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.85	(0.44)	1.48 - 3.38	4.03	(0.40)	3.52 - 5.00	2.79	(0.33)	2.33 - 3.41	4.03	(0.45)	3.50 - 4.95
Distinctiveness	3.33	(0.48)	2.80 - 4.80	3.39	(0.37)	2.85 - 4.35	2.31	(0.37)	1.39 - 2.70	2.23	(0.26)	1.60 - 2.55
Frequency	2.49	(0.58)	1.68 - 3.95	2.51	(0.47)	1.53 - 3.15	2.46	(0.47)	1.85 - 3.45	2.50	(0.65)	1.38 - 3.80
Likeness	3.84	(0.32)	3.05 - 4.62	3.90	(0.40)	3.05 - 4.62	3.83	(0.41)	2.48 - 4.29	3.83	(0.52)	2.75 - 4.71

Stimuli were presented in eight discrete blocks, however in order not to disrupt performance, only four were discernible. Items were blocked to reduce the likelihood of associated famous faces appearing in succession and possibly causing associative priming. Stimuli were randomized within blocks and blocks were rotated between participants. During the breaks between the four blocks, participants were reminded to "respond as quickly and accurately as possible".

The experiment commenced with 20 practice trials (10 famous and 10 unfamiliar) and the presentation of experimental trials was directly preceded by four lead-in trials as means of concealing initial hesitations. Both practice and lead-in trials contained a representative sample of stimuli (early and late acquired, typical and distinctive, famous and unfamiliar, male and female) to preclude the possibility of bias effects. However, none of these items were repeated as experimental trials or analysed. The task took approximately 15 minutes.

3.1.3 Results

Prior to the analysis, four participants were replaced because either they failed to recognise at least 75% of the famous faces, or failed to recognise 75% of the famous and unfamiliar faces when pooled together. The first criterion was implemented to ensure that the latency analysis is based on a sufficient number of responses and the second, to prevent inclusion of data derived from adverse response strategies.

The accuracy analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions; b) reaction times deemed outliers. As a result of incorrect familiarity decisions made to famous faces, 414 (12.7%) of the responses were removed. Forty five (1.4%) outliers were removed for being slower than 3 standard deviations away from the mean or faster than 300ms. However, the accuracy analysis should be addressed with caution as it is impossible to distinguish between a genuine error and unfamiliarity with a target face, thus limiting its validity.

The latency analysis was carried out on correct responses only. Table 3.2 shows the mean correct response latencies in each condition of the experiment and percentage error rates. Unfamiliar face summary data is also reported. For inspection purposes, '% misclassifications', referring to the percentage of incorrect familiarity decisions, are documented.

Table 3.2. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Experiment 1).*

	Early Acquired		Late Acquired		Unfamiliar
	Distinctive	Typical	Distinctive	Typical	
Mean RT	730	754	750	775	768
SD	(121)	(111)	(121)	(120)	(123)
% Error	7.11	14.58	14.46	20.10	6.46
% Misclass.	6.00	13.36	12.87	18.50	4.63

Reaction Times. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of AoA was significant in the by-subject analysis, $F_1(1,23) = 13.98$, MSe = 696.79, $p = .001$, reflecting the faster recognition of early faces (742 ms) compared with those acquired later (762 ms). However, AoA was not significant in the by-items analysis; $F_2(1,132) = 2.82$, MSe = 9,308.85, $p = .10$. The main effect of distinctiveness reached significance in both analyses, $F_1(1,23) = 7.49$, MSe = 1,867.40, $p = .012$; $F_2(1,132) = 3.92$, MSe = 9,308.85, $p = .05$, maintaining the customary advantage found for distinctive faces (740 ms) compared with typical faces (764 ms). The interaction between AoA and distinctiveness was not significant in either analysis, $F_1(1,23) = 0.004$, MSe = 1,054.27, $p = .95$; $F_2(1,132) = 0.03$, MSe = 9,308.85, $p = .86$. Inspection of the data in Table 3.2 suggested that AoA and distinctiveness are perfectly additive.

Errors. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. The main effect of AoA was significant in both analyses, $F_1(1,23) = 23.09$, MSe = 4.94, $p<.001$; $F_2(1,132) = 4.29$, MSe = 18.89, $p<05$, stemming from the more accurate recognition of early acquired faces (mean n = 3.7, 10.9%) compared with those late acquired (mean n = 5.9, 17.3%). In addition, the main effect of distinctiveness was significant in both analyses, $F_1(1,23) = 30.15$, MSe = 3.96, $p<.001$; $F_2(1,132) = 4.46$, MSe = 18.89, $p<.05$, confirming the tendency for distinctive faces (mean n = 3.7, 10.8%) to be recognised more accurately than typical faces (mean n = 5.9, 17.3%). The interaction was not significant in either analysis, $F_1(1,23) = .93$, MSe = 2.52, $p = .35$; $F_2(1,132) = 0.09$, MSe = 18.89, $p = .77$, signalling additivity.

A glance at the raw data showed that particular target faces attracted a lot of errors, thus compromising the validity of the latency and accuracy analyses: the latency data could include attempts to recognise faces which are not particularly familiar and an incorrect response could be genuine error or reflect unfamiliarity with the target face. Therefore, an additional analysis was carried out using only those stimuli recognized by at least 75% of participants. The latency and accuracy data were submitted to by-subjects repeated measures ANOVAs with AoA and distinctiveness as factors. In the latency analysis, the main effect of AoA was significant, assuming the expected trend, $F1(1,23) = 8.42$, MSe = 791.56, $p<.01$. The main effect of distinctiveness approached significance and again the trend conformed to expectation, $F1(1,23) = 3.74$, MSe = 1,256.37, $p = .07$. Of particular note, the interaction reached borderline significance, $F1(1,23) = 4.09$, MSe = 850.54, $p = .06$. Tukey tests ($\alpha = .05$) revealed that distinctiveness exclusively modulated familiarity decision performance associated with late acquired faces, and AoA effects were restricted to the recognition of typical faces (HSD = 23). Thus, the interaction took the predicted form. After the accuracy data had been arcsine transformed (percentages were submitted to the ANOVA due to differences in set size), AoA, $F1(1,23) = 25.43$, MSe = 6.95E-03, $p<.001$, and distinctiveness, $F1(1,23) = 47.71$, MSe = 4.58E-03, $p<.001$, attained significance, assuming expected trends. The interaction failed to attain significance, $F1(1,23) = 2.37$, MSe = 3.25E-03, $p = .14$. Independent t-tests confirmed that the elimination of poorly recognised famous faces failed to distort set matching: the sets remained matched for frequency and likeness; the two distinctive and two typical sets remained matched by distinctiveness; the two early sets and two late sets remained matched by AoA (in each case $t \leq 1.37$).

3.1.4 Discussion

Experiment 1 found customary AoA and distinctiveness effects in both the latency and accuracy analyses, but the interaction failed to reach significance in either analysis: evidently speed-accuracy trade-offs were not in operation. The relationship between AoA and distinctiveness was characterised by more or less perfect additivity. It is interesting to note that the reaction times to the famous faces in this experiment (752 ms) were faster than those in the Pilot Experiment (885 ms) and that twice the number of errors were made to famous faces in this experiment (14.1 %) compared with the Pilot (6.4 %). This would suggest that participants are trading accuracy for speed. However, an inspection of the errors made to unfamiliar faces in this experiment (6.5 %) and the Pilot Experiment (4.9 %) indicates that this is not likely to be the case. Indeed, a glance at the raw data demonstrated that particular target faces attracted a disproportionately large number of errors. Rather than stemming from the misclassification of famous faces, the reduced accuracy may have arisen because certain stimuli were insufficiently familiar to participants. If this is the case, then noise will not only be added to the accuracy analysis but the latency analysis will capture attempts to recognise faces which are not sufficiently familiar.

The analysis carried out on the reduced stimulus set revealed that the significant effects of AoA and distinctiveness on familiarity decisions in the latency and accuracy analyses had been maintained. Interestingly, an interaction of borderline significance emerged in the latency analysis taking the expected form: AoA effects were restricted to the processing of typical faces and distinctiveness affected the recognition of late acquired faces. The interaction in the accuracy analysis failed to attain significance, but this is not surprising given that the reduced experimental sets had been refined on the basis of recognition accuracy.

Although the stimulus set sizes were rendered uneven after the extraction of stimuli, statistical set matching had not been compromised. Nevertheless, the performance latency of the sets from which items had been eliminated (early acquired typical, late acquired typical, and late acquired distinctive), may have been artificially improved relative to the early acquired distinctive set, from which no stimuli had been removed. If performance to faces in these three sets had been improved by a constant amount, then the outcome would be an interaction akin to that predicted. Even with the best intentions, post-hoc tweaking of the data is dubious.

3.2 Experiment 2

3.2.1 Introduction

Investigation into the relationship between AoA and distinctiveness was pursued once again with up-to-date images of famous faces. This time faces were rated for familiarity so that the database could be screened to ensure the removal of faces insufficiently familiar to participants. Faces which had attracted a lot of errors in Experiment 1 were also discarded from the database.

Once again, a familiarity decision task was administered and performance with early acquired distinctive faces, early acquired typical faces, late acquired distinctive faces and late acquired typical faces was compared. Effects of AoA and distinctiveness were expected and qualification of these effects by the hypothesised interaction was predicted.

3.2.2 Method

Participants. Twenty three undergraduate students from the University of York (18 female, 5 male) participated in this experiment in return for course credit or £2, so long as they were aged between 18 – 21 years old and had lived in the UK for at least 18 years. The mean age of participants was 18.96 years old (SD = 0.64), and ages ranged from 18-21 years old.

Design. This 2 x 2 within-subjects design constituted a factorial manipulation of AoA (with two levels, early and late) and distinctiveness (with two levels, distinctive and typical). This led to the formation of four sets: early acquired distinctive; early acquired typical; late acquired distinctive; late acquired typical. A face familiarity decision task yielded two dependent variables; response latency and response accuracy.

Materials. Before creating the stimulus set, a number of issues associated with the database required resolution. Firstly, from the results of Experiment 1 it was evident that a subset of the famous faces was not sufficiently familiar to participants. At this point in proceedings, the database was rated for familiarity and faces were removed if they were not rated as 'familiar' at the very least. In addition, items that had received error rates in excess of 20% during Experiment 1 were discarded from the database.

Secondly, problems associated with the creation of the database, namely the initial difficulty obtaining early acquired faces which were also rated as typical, could be symptomatic of problems with the distinctiveness ratings. Moore and Valentine (1998) reported a significant correlation between AoA and distinctiveness indicating that faces acquired early tended to be rated as more distinctive than late acquired faces, $r(99) = -.37$, $p<.01$. However, the same relationship investigated with the current database failed to reach significance, $r(258) = -.10$, $p>.05$). This was not surprising given the efforts invested to increase the number of early acquired typical faces in the database.

The effort to increase the number of early acquired typical faces may have masked a difference in the way in which early and late acquired faces were rated for distinctiveness. Thus, it was necessary to check the validity of the distinctiveness ratings as a prerequisite to further testing. It is possible that undergraduates overestimated the distinctiveness of early acquired faces for the reasons posited by the hypothesis under scrutiny: if a 'uniform' face-space is adopted to account for the childhood attenuation of distinctiveness effects then early acquired faces could be encoded as distinctive. Undergraduate ratings of distinctiveness were compared with distinctiveness ratings made by 22 adults (mean age = 43.60, S.D. = 9.49), for whom the 'early acquired' faces would be late acquired. The instruction booklets and rating scales were identical to those given to the undergraduates (see p. 76). A significant correlation was found between undergraduate and adult ratings of distinctiveness, $r(258) = .80$, $p<.01$, indicating that distinctiveness ratings were not affected by AoA. Moreover, a significant correlation was found between undergraduate and adult ratings of early acquired faces alone, $r(128) = .85$, $p<.01$. This was marginally stronger than the corresponding late acquired correlations, $r(128) = .74$, $p<.01$. Undergraduates appear to be rating facial distinctiveness in the same way as adults, even faces early acquired by one cohort and late acquired by the other.

It is possible that ratings of distinctiveness inadvertently tapped familiarity. The distinctiveness of early acquired faces could be overestimated on the premise that they have been known for longer than late acquired faces and, therefore, have a higher perceived familiarity. However, the relationship between distinctiveness and familiarity was weak, $r(258) = .25$, $p<.01$, indicating a negligible effect of familiarity on distinctiveness. The undergraduate distinctiveness ratings continued to be used.

The stimulus set comprised 72 famous faces taken from the current database (18 early acquired distinctive, 18 early acquired typical, 18 late acquired distinctive, 18 late acquired typical). Independent t-tests confirmed that the sets were matched for frequency and likeness and that the two distinctive and two typical sets were each matched for distinctiveness and that the two early sets and two late sets were each matched for AoA (in each case $t \leq 1.54$). See Table 3.3 for summary data pertaining to the creation of experimental sets.

Famous faces were matched with an equal number of unfamiliar faces by gender, race and approximate age. The unfamiliar faces received the same digital treatment as the famous faces.

Procedure and Apparatus. The familiarity decision task was run on PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to the images, which were presented in the centre of the screen. Each experimental trial began with a 500 ms delay prior to image onset. Faces were displayed until a response was made. Participants logged 'familiar' and 'unfamiliar' responses by pressing the 'z' and '/' keys on the keyboard with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 and their dimensions subtended 70 x 150 mm approximately. Participants viewed these from a distance of approximately 30cm.

Stimuli were presented in two discrete blocks, however owing to the short duration of the experiment, this was not discernible. This subdivision of blocks minimised the likelihood of associative priming. Stimuli were randomized within blocks and blocks were counter-balanced.

The experiment commenced with 20 practice trials (10 famous and 10 unfamiliar) and presentation of the experimental trials was directly preceded by four lead-in trials. Both practice and lead-in trials contained a representative sample of stimuli (early and late acquired, typical and distinctive, famous and unfamiliar, male and female) to preclude the possibility of bias effects. None of these items were repeated as experimental trials or analysed. The task took approximately 10 minutes.

Table 3.3. *Mean Ratings as a Function of Experimental Set (Experiment 2).*

	Early acq. & distinctive			Late acq. & distinctive			Early acq. & typical			Late acq. & typical		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.67	(0.31)	2.14 - 3.24	3.93	(0.44)	3.48 - 5.00	2.86	(0.43)	2.00 - 3.43	4.14	(0.47)	3.52 - 4.95
Distinctiveness	3.50	(0.49)	2.90 - 4.80	3.36	(0.49)	2.75 - 4.35	2.23	(0.36)	1.65 - 2.65	2.36	(0.27)	1.80 - 2.70
Frequency	2.42	(0.67)	1.18 - 3.95	2.44	(0.39)	1.70 - 3.10	2.35	(0.62)	1.32 - 3.45	2.61	(0.41)	1.43 - 3.05
Likeness	3.92	(0.49)	3.05 - 4.67	3.95	(0.32)	3.25 - 4.48	3.91	(0.38)	3.00 - 4.33	3.80	(0.45)	2.90 - 4.71

3.2.3 Results

Prior to the analysis, eight participants were replaced for failing to recognise at least 75% of the famous faces and not recognising 75% of the famous and unfamiliar faces when pooled together. The first criterion was implemented to ensure that the latency analysis is based on a sufficient number of responses and the second, to prevent inclusion of data derived from adverse response strategies.

The accuracy analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions; b) reaction times deemed outliers. As a result of incorrect familiarity decisions made to famous faces, 191 (11.5%) of the responses were removed. Twenty four (1.4%) outliers were removed for being slower than 3 standard deviations away from the mean or faster than 300ms. Again, the accuracy analysis should be addressed with caution as it is impossible to distinguish between a genuine error and unfamiliarity with a target face, thus limiting its validity.

The latency analysis was carried out on correct responses only. Table 3.4 shows the mean correct response latencies in each condition of the experiment and percentage error rates. Unfamiliar face summary data is also reported. For inspection purposes, '% misclassifications', referring to the percentage of incorrect familiarity decisions, are documented.

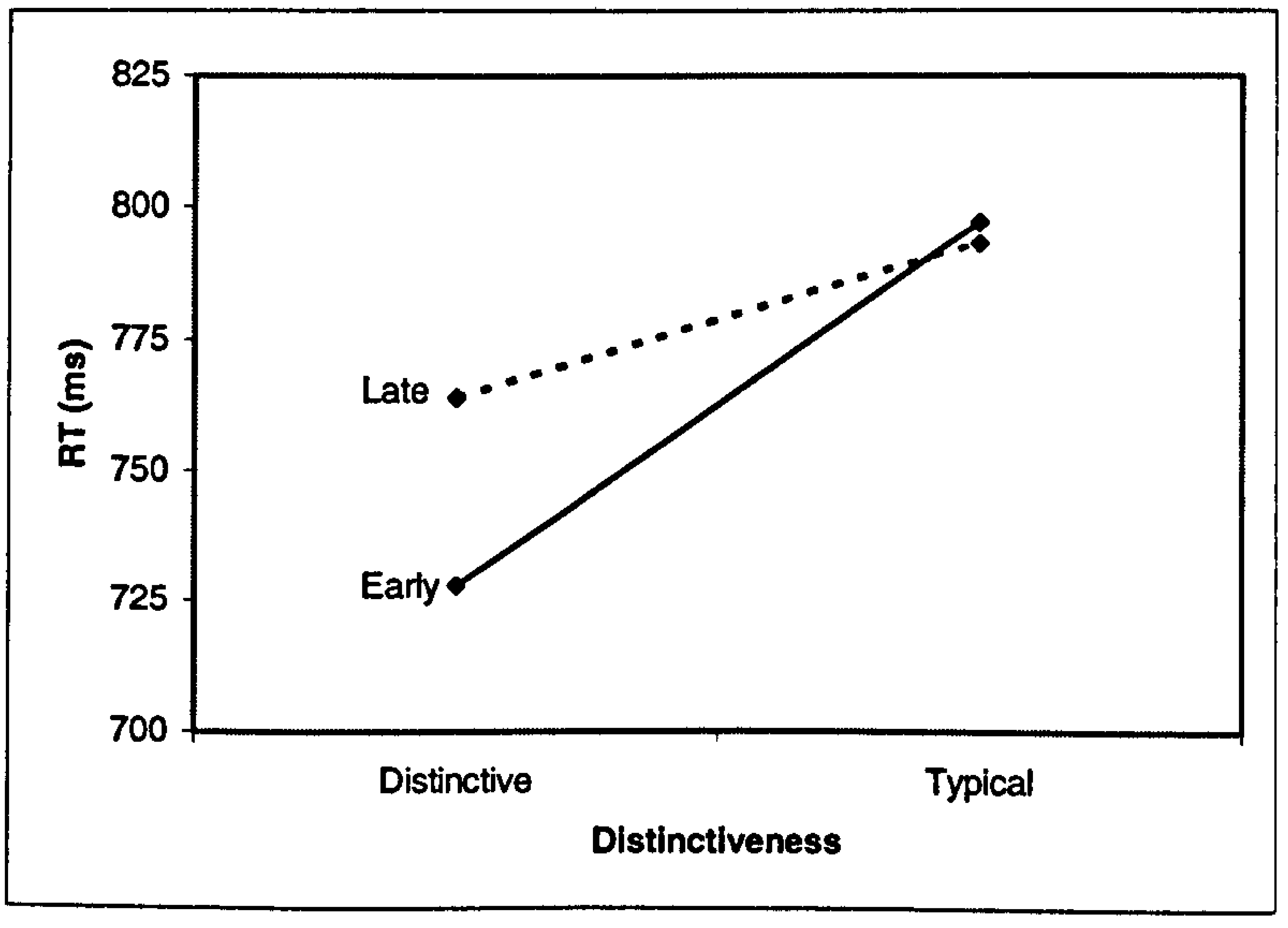
Table 3.4. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Experiment 2).*

	Early Acquired		Late Acquired		Unfamiliar
	Distinctive	Typical	Distinctive	Typical	
Mean RT	728	797	764	793	791
SD	(99)	(123)	(110)	(105)	(115)
% Error	9.42	13.53	8.94	20.05	5.68
% Misclass.	8.45	12.32	7.25	18.12	4.05

Reaction Times. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The by-subjects analysis yielded a significant main effect of AoA, $F_1(1,22) = 7.72$, MSe = 783.54, $p<.001$, with early acquired faces (763 ms) recognised faster than late acquired faces (779 ms). This effect was not significant by-items, $F_2(1,68) = 0.79$, MSe = 5,884.93, $p = .38$. The main effect of distinctiveness was significant in both analyses, $F_1(1,22) = 31.42$, MSe = 1,763.72, $p<.001$; $F_2(1,68) = 7.28$, MSe = 5,884.93, $p<.01$, stemming from the faster responses made to distinctive faces (746 ms) compared with typical faces (795 ms).

The two-way interaction between AoA and distinctiveness was significant in the by-subject analysis alone, $F_1(1,22) = 5.75$, MSe = 1,650.31, $p<.05$; $F_2(1,68) = 0.76$, MSe = 5,884.93, $p = .39$, see Figure 3.1. With alpha set at .05, *a posteriori* Tukey comparisons (HSD = 33) revealed that the customary AoA advantage was restricted to distinctive faces, leaving typical faces unaffected by AoA. In addition, distinctiveness affected the recognition of early acquired faces, but not those acquired late.

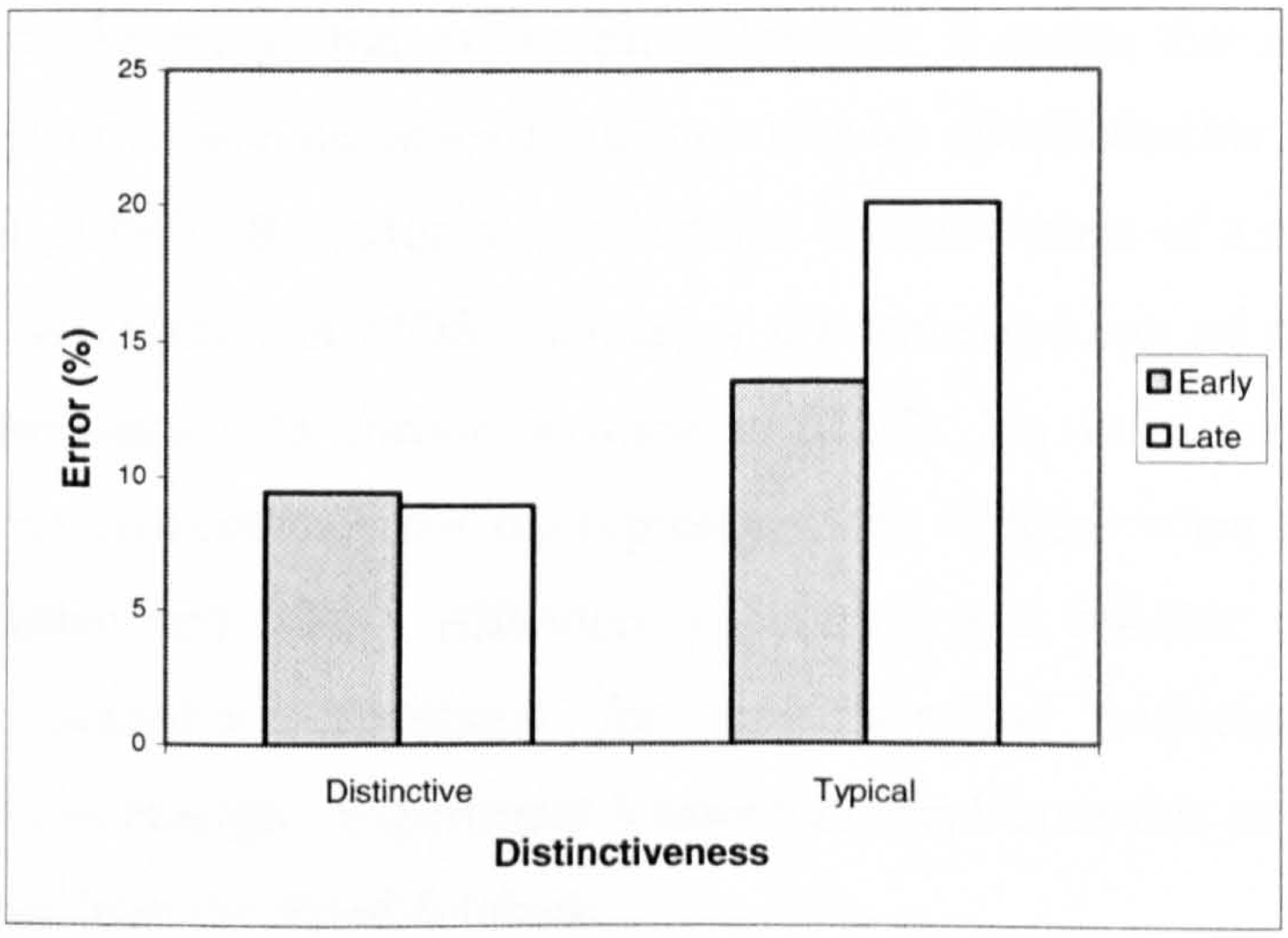
Figure 3.1. *Effect of AoA on Familiarity Decision RTs as a Function of Distinctiveness (Experiment 2).*



Errors. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. The by-subjects analysis yielded a significant main effect of AoA, $F_1(1,22) = 4.47$, MSe = 1.52, $p<.05$, with early acquired faces (mean n = 2.1, 11.5%) recognised more accurately than those late acquired (mean n = 2.6, 14.5%). This effect was not reproduced by-items, $F_2(1,68) = 0.98$, MSe = 8.85, $p = .33$. The main effect of distinctiveness reached significance in both analyses, $F_1(1,22) = 11.85$, MSe = 3.64, $p<.01$; $F_2(1,68) = 6.23$, MSe = 8.85, $p<.05$, confirming the observed tendency for distinctive faces (mean n = 1.7, 9.2%) to be recognised more accurately than typical faces (mean n = 3.0, 16.8%).

A significant two-way interaction emerged between AoA and distinctiveness in the by-subjects analysis, $F_1(1,22) = 8.34$, MSe = 1.10, $p<.01$, (see Figure 3.2), but not in the by-items analysis, $F_2(1,68) = 1.32$, MSe = 8.85, $p = .26$. Tukey tests confirmed that typical faces were affected by AoA, whereas distinctive faces were unaffected. Also, distinctiveness affected the recognition accuracy of late acquired faces, but not those acquired early (HSD = 0.86). Whilst its form mirrored that hypothesized, the interaction diametrically opposed that found in the latency analysis. However, genuine errors cannot be discriminated from unfamiliarity with a face, possibly limiting the credibility of this result.

Figure 3.2. *Effect of AoA on Familiarity Decision Accuracy as a Function of Distinctiveness (Experiment 2).*



3.2.4 Discussion

Familiarity decision speed was affected by AoA and distinctiveness with effects conforming to expectation. A significant interaction emerged but it countered predictions: AoA affected performance associated with distinctive faces and distinctiveness effects were restricted to early acquired faces. The accuracy analysis also recovered effects of AoA and distinctiveness. However, these effects were qualified by an interaction diametrically opposed to that found in the latency analysis: AoA affected the accuracy of typical face familiarity decisions and distinctiveness influenced performance associated with late acquired faces. Although this interaction mirrors predictions, it does not constitute particularly compelling evidence, owing to the problems discriminating unfamiliarity with a target face from genuine misclassifications. However, the accuracy analysis served to demonstrate the speed-accuracy trade-offs were not in operation.

The unexpected interaction in the latency analysis stemmed from the particularly fast responses made to faces in the late acquired typical set. An inspection of the mean ratings (Table 3.3) demonstrated that this set contained more frequently encountered faces than the other sets. In fact, the frequency ratings of the late acquired typical and early acquired typical sets are the most poorly matched pair of ratings in this experiment. The interaction may be an artefact of the inflated frequency ratings in this set.

The influence of age transformations on the encoding and storage of representations for recognition is unclear. However, it seems that increasing the perceived age of a face increases its distinctiveness (Deffenbacher et al., 1998; O'Toole et al., 1997). By virtue of the fact that distinctiveness of a representation is coded by its location in MDS, increasing the perceived age of a face should cause a representation to change location in MDS. By employing up-to-date images, this experiment may not tap representations of faces when first acquired and first encoded into MDS. Although speculative, it is possible that this may impede the sensitivity necessary for an interaction between AoA and distinctiveness to emerge. Experiment 3 essentially replicates this experiment, but recruits images from the dated database.

3.3 Experiment 3

3.3.1 Introduction

As the final experiment to explore the mutual influence of AoA and distinctiveness on familiarity decisions, Experiment 3 replicated Experiment 2 with dated images. Familiarity decisions were made to early acquired distinctive faces, early acquired typical faces, late acquired distinctive faces and late acquired typical faces. As before, effects of AoA and distinctiveness were predicted in conjunction with the hypothesised interaction between the variables.

3.3.2 Method

Participants. Forty nine undergraduate students from the University of York (39 female, 10 male) participated in this experiment, in return for course credit or a payment of £2. All participants had lived in the UK for at least 18 years and the age range tested constituted 18 – 22 years old. The mean age of participants was 19.65 years old (SD = 0.83)[15].

Design. This 2 x 2 within-subjects design constituted a factorial manipulation of AoA (with two levels, early and late) and distinctiveness (with two levels, distinctive and typical). This led to the formation of four sets: early acquired distinctive; early acquired typical; late acquired distinctive; late acquired typical. A face familiarity decision task yielded two dependent variables; response latency and response accuracy.

Materials. The stimulus set comprised a subset of 72 images (18 early acquired distinctive, 18 early acquired typical, 18 late acquired distinctive, 18 late

[15] A large number of participants were tested to maximise the opportunity for an interaction to emerge. In a preliminary analysis, based on the results of 22 participants, the predicted interaction was recovered but it fell just short of significance. In addition, it became apparent that the majority of participants failing to meet the inclusion criteria were towards the bottom of the age range. As a consequence, the age range was extended to encompass 22 year olds: the database had been created nearly a year prior to this experiment, thus it had been rated by participants, now 19 – 22 years old.

acquired typical). If a face had been acquired since 1996, then images from the current database were used to depict these late acquired items: age-related change was not deemed sufficiently dramatic to warrant dated images. Stimuli acquired pre-1996 were taken from the dated database. Eighty three percent of the stimulus set captured dated images. Of the remaining images from the current database, half were late acquired and distinctive and half late acquired and typical. Seventy eight percent of the celebrities featured were also included in the stimulus set featured in Experiment 2, thus maximising stimulus overlap and facilitating comparisons[16].

Independent t-tests confirmed that the four sets were matched for frequency and likeness and that the two distinctive and two typical sets were each matched for distinctiveness and that the two early sets and two late sets were each matched for AoA (in each case $t \leq 1.69$). See Table 3.5 for summary data pertaining to the creation of experimental sets.

Famous faces were matched with an equal number of unfamiliar faces using gender, race and approximate age as controls. These unfamiliar faces were digitally treated in the same manner as the famous faces.

Procedure and Apparatus. The familiarity decision task was run on PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to the images, which were presented in the centre of the screen. Participants were warned that some of the images may be dated depictions of particular faces and informed that this should not affect their decisions. Each experimental trial began with a 500 ms delay prior to image onset. Faces were displayed until a response was made. Participants logged 'familiar' and 'unfamiliar' responses by pressing the 'z' and '/' keys on the keyboard with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 and their dimensions subtended 70 x 150 mm approximately. Participants viewed these from a distance of approximately 30cm.

---

[16] Included in the 78% overlap between experiments are identical current images of the same celebrity. Therefore, combined statistical analyses are not appropriate because some of the stimuli occurred in both experiments.

Stimuli were presented in two discrete blocks, however owing to the short duration of the experiment this was not discernible. This subdivision of blocks minimised the likelihood of associative priming. Stimuli were randomized within blocks and blocks were counter-balanced.

The experiment began with 20 practice trials (10 famous and 10 unfamiliar) and presentation of the experimental trials was directly preceded by four lead-in trials. Both practice and lead-in trials contained a representative sample of stimuli (early and late acquired, typical and distinctive, famous and unfamiliar, male and female) to preclude the possibility of bias effects. None of these items were repeated as experimental trials or analysed. The task took approximately 10 minutes.

Table 3.5. *Mean Ratings as a Function of Experimental Set (Experiment 3).*

	Early acq. & distinctive			Late acq. & distinctive			Early acq. & typical			Late acq. & typical		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.87	(0.38)	2.14 - 3.38	3.94	(0.44)	3.48 - 4.71	2.85	(0.42)	2.29 - 3.43	4.08	(0.43)	3.52 - 4.95
Distinctiveness	3.21	(0.32)	2.77 - 3.74	3.31	(0.54)	2.75 - 4.35	2.29	(0.28)	1.70 - 2.70	2.36	(0.24)	1.88 - 2.70
Frequency	2.56	(0.46)	2.00 - 3.95	2.58	(0.63)	1.70 - 4.00	2.33	(0.66)	1.21 - 3.45	2.61	(0.42)	1.43 - 3.05
Likeness	3.57	(0.65)	2.16 - 4.69	3.80	(0.40)	3.00 - 4.48	3.51	(0.65)	1.42 - 4.48	3.51	(0.60)	2.36 - 4.24

3.3.3 Results

Prior to the analysis, 25 participants were replaced for failing to recognise at least 75% of the famous faces, and not recognising 75% of the famous and unfamiliar faces when pooled together[17]. The first criterion was implemented to ensure that the latency analysis is based on a sufficient number of responses and the second, to prevent inclusion of data derived from adverse response strategies.

The accuracy analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions; b) reaction times deemed outliers. As a result of incorrect familiarity decisions made to famous faces, 426 (12.1%) of the responses were removed and 41 (1.2%) outliers were removed for being slower than 3 standard deviations away from the mean or faster than 300ms. Again, the accuracy analysis should be addressed with caution as it is impossible to distinguish between a genuine error and unfamiliarity with a target face, thus limiting its validity.

The latency analysis was carried out on correct responses only. Table 3.6 shows the mean correct response latencies in each condition of the experiment and percentage error rates. Unfamiliar face summary data is also reported. For inspection purposes, '% misclassifications', referring to the percentage of incorrect familiarity decisions, are documented.

---

[17] Presumably such a large number of participants failed to meet the recognition criteria because of the difficulty associated with recognising dated images of famous faces.

Table 3.6. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Experiment 3).*

	Early Acquired		Late Acquired		Unfamiliar
	Distinctive	Typical	Distinctive	Typical	
Mean RT	698	724	709	750	735
SD	(120)	(131)	(112)	(121)	(133)
% Error	8.28	15.99	12.02	16.78	6.33
% Misclass.	7.48	14.85	11.00	15.19	4.29

Reaction Times. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of AoA was significant in the by-subjects analysis, $F_1(1,48) = 7.71$, MSe = 2,103.15, $p<.01$, with early acquired faces (711 ms) recognised more quickly than late acquired faces (730 ms). This effect was not significant in the by-items analysis, $F_2(1,68) = 1.50$, MSe = 5,541.18, $p = .23$. The main effect of distinctiveness was significant in both analyses, $F_1(1,48) = 37.68$, MSe = 1,495.06, $p<.001$; $F_2(1,68) = 4.48$, MSe = 5,541.18, $p<.05$, reflecting the faster recognition of distinctive faces (703 ms) compared with typical items (737 ms). The interaction between AoA and distinctiveness failed to reach significance in either analysis, $F_1(1,48) = 1.37$, MSe = 2,243.73, $p = .25$; $F_2(1,68) = 0.16$, MSe = 5,541.18, $p = .69$, despite exhibiting tendencies towards the predicted form.

Errors. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. The main effect of AoA approached significance in the by-subjects analysis, $F_1(1,48) = 3.87$, MSe = 2.11, $p = .06$, with early acquired faces (mean n = 2.2, 12.2%) recognised more accurately than late acquired faces (mean n = 2.6, 14.4%). This finding was not replicated in the by-items analysis, $F_2(1,68) = 0.40$, MSe = 56.09, $p = .53$. The main effect of distinctiveness was significant in the by-subjects analysis, $F_1(1,48) = 27.00$, MSe = 2.29, $p<.001$, with distinctive faces (mean n = 1.8, 10.2%) recognised more accurately than typical faces (mean n = 3.0, 16.4%). Distinctiveness was not significant by-items, $F_2(1,68) = 3.00$, MSe = 56.09, $p = .09$. The interaction failed to reach significance, $F_1(1,48) = 1.99$, MSe = 1.73, $p = .16$; $F_2(1,68) = 0.17$, MSe = 56.09, $p = .68$.

3.3.4 Discussion

Experiment 3 yielded effects of AoA and distinctiveness in the latency analysis. The interaction exhibited tendencies towards the predicted form, falling short of significance despite testing a large number of participants: distinctive faces showed an AoA effect of only 11 ms compared with the 26 ms difference between early acquired and late acquired typical faces; early acquired faces showed a distinctiveness effect of 26 ms whereas late acquired distinctive faces were recognised on average 41 ms faster than late acquired typical faces. Recognition accuracy was affected by distinctiveness, but not AoA, and the interaction was not significant. Speed-accuracy trade-offs did not affect results. Even when noise through age-related changes in appearance since acquisition have been reduced, the relationship appears to be additive.

Whilst, mean recognition accuracy for famous faces in Experiment 3 (13.3%) is similar to that found in Experiment 2 (13.0%), reaction times in Experiment 3 (720 ms) are faster than in Experiment 2 (771 ms). One would have expected dated images to be harder to recognise than current images, incurring processing costs as a consequence. Moreover, one would have expected this effect to be particularly pronounced in the early acquired condition where faces have undergone greater age-related changes. However, the difference in reaction times between the experiments was identical for the early and late sets. This may have arisen because present task demands were harder than those imposed by Experiment 2 and so participants achieving the inclusion criteria may have possessed a much higher level of familiarity with the faces than participants fulfilling Experiment 2 criteria.

Of note, the mean frequency rating (see Table 3.5) of the late acquired typical set is elevated relative to the early acquired typical set, as it is in Experiment 2. Recall that the unexpected interaction in Experiment 2 was attributed to the elevated mean frequency of the late acquired typical faces. The poor matching of the early and late acquired typical sets appears not to compromise performance in this Experiment, and therefore unlikely to be the cause of the unexpected interaction in Experiment 2.

3.4 General Discussion: Experiments 1 to 3

Experiments 1 to 3 sought to ascertain the nature of the relationship between AoA and distinctiveness to offer a refinement of the mechanisms proposed to underpin the variables. However, the findings lacked coherence. To aid interpretation and bypass extraneous detail, accuracy analyses are not subject to discussion except when they have a direct bearing on reaction times. Moreover, the validity of the accuracy analyses is limited because misclassifications stemming from unfamiliarity with a target face cannot be discriminated from genuine errors. This commonly encountered problem often causes face recognition studies to neglect accuracy analyses altogether, enlisting the data for inspection purposes only (e.g. Bruce et al., 1993b). Inspection of the accuracy data indicated that speed-accuracy trade-offs were not operating to compromise results in the latency analysis.

The lack of coherence in this series of experiments makes it difficult to disconfirm the hypothesis that AoA and distinctiveness interact. However, the hypothesis has certainly not been confirmed. In those instances where the predicted interaction arose, the experiments were not without problems. The Pilot Experiment and the reduced dataset of Experiment 1 yielded the predicted interaction: AoA effects were restricted to the recognition of typical faces and distinctiveness exclusively modulated responses to late acquired faces. However, in the Pilot Experiment, performance accuracy was conspicuously lower for distinctive early acquired faces compared with distinctive late acquired faces. This flagged the possibility that interactions in the latency and accuracy analyses could be artefactual. As age of image had not been standardised, it was possible that early acquired sets comprised more dated images than late acquired sets. In the reduced dataset of Experiment 1, poorly recognised target faces had been removed. However, post-hoc tweaking of data is potentially dubious even when stimuli are removed in accordance with a pre-defined error threshold.

The predicted interaction was not recovered from Experiments 1 to 3. Experiment 1 found additive effects of AoA and distinctiveness and Experiment 2 displayed interactive effects which countered those predicted: AoA effects were restricted to the recognition of distinctive faces and distinctiveness exclusively affected performance associated with early acquired faces. The interaction in Experiment 3 failed to reach significance and only exhibited tendencies towards the predicted form, even after testing 49 participants. A definitive interpretation of these results is not possible, but the inclinations towards additivity are more persuasive than various interactions recovered under suboptimal circumstances.

Compatible with additivity in the behavioural data, ratings of AoA and distinctiveness were not correlated. Had an interaction between AoA and distinctiveness emerged, it would have been conceivable that the variables would also be correlated[18]: Reflecting the behavioural hypothesis, it is possible that distinctiveness ratings could have been made with recourse to the temporally-defined layers of MDS hypothesised. However, on the basis of the present findings there is no evidence to suggest that AoA and distinctiveness are integrated or correlated. This indicates that the correlation found previously in the literature (Moore & Valentine, 1998) may have been an artefact of inadequate distinctiveness ratings which tapped the distinctiveness of the person rather than the face.

Whilst it is tempting to think that this null correlation bolsters the contention that the behavioural data be best described as additive, an interaction in the behavioural data will not necessarily be reciprocated by a correlation. Provided that participants can rate the distinctiveness of early and late acquired faces with recourse to the population of faces experienced over a lifetime, regardless of

[18] This correlation could be positive or negative depending on the version of child face-space adopted. In H. Ellis' (1992) uniform model of child space, characterised by parameters akin to those structuring adult face-space coupled with an attenuated density gradient, all representations are distinctive. In this case, AoA and distinctiveness would be negatively correlated. In H. Ellis' differential model, smaller than adult space but with a preserved distribution of typical and distinctive faces, AoA and distinctiveness would be positively correlated.

whether or not underlying mechanisms are integrated, the variables will appear not to be related. Moreover, ratings are not totally objective and should not serve as a proxy for performance in behavioural studies. For example, Wickham et al., (2000) found that distinctiveness rating instructions which emphasised the distinctiveness of a face among other faces, predicted hits better than instructions which emphasised deviation from typicality. Cited by Wickham et al., Vokey and Read (1992) found that the latter instructions aptly predicted false positives. Thus, correlations by no means prop experimental findings.

Additivity in the behavioural data indicates that AoA and distinctiveness are sub-served by independent mechanisms. Couched in terms of A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis, the loss of network plasticity hypothesised to underpin AoA effects influences all representations equally, regardless of location in MDS. Additionally, there is no evidence to suggest that faces are stored with reference to the MDS into which they were initially encoded. As Valentine (1991) proposed, the quality of representations evolves with the emergence of a normal distribution of representations over time. Figure 3.3 charts the evolution of the uniform child face-space into the centrally clustered adult space illustrating how an early acquired face, once seen as distinctive, becomes typical as an increasing number of neighbouring representations are assimilated into MDS. There is no evidence to suggest that A. Ellis and Lambon Ralph's neural plasticity hypothesis and Valentine's MDS require elaboration and integration.

[Original in colour]

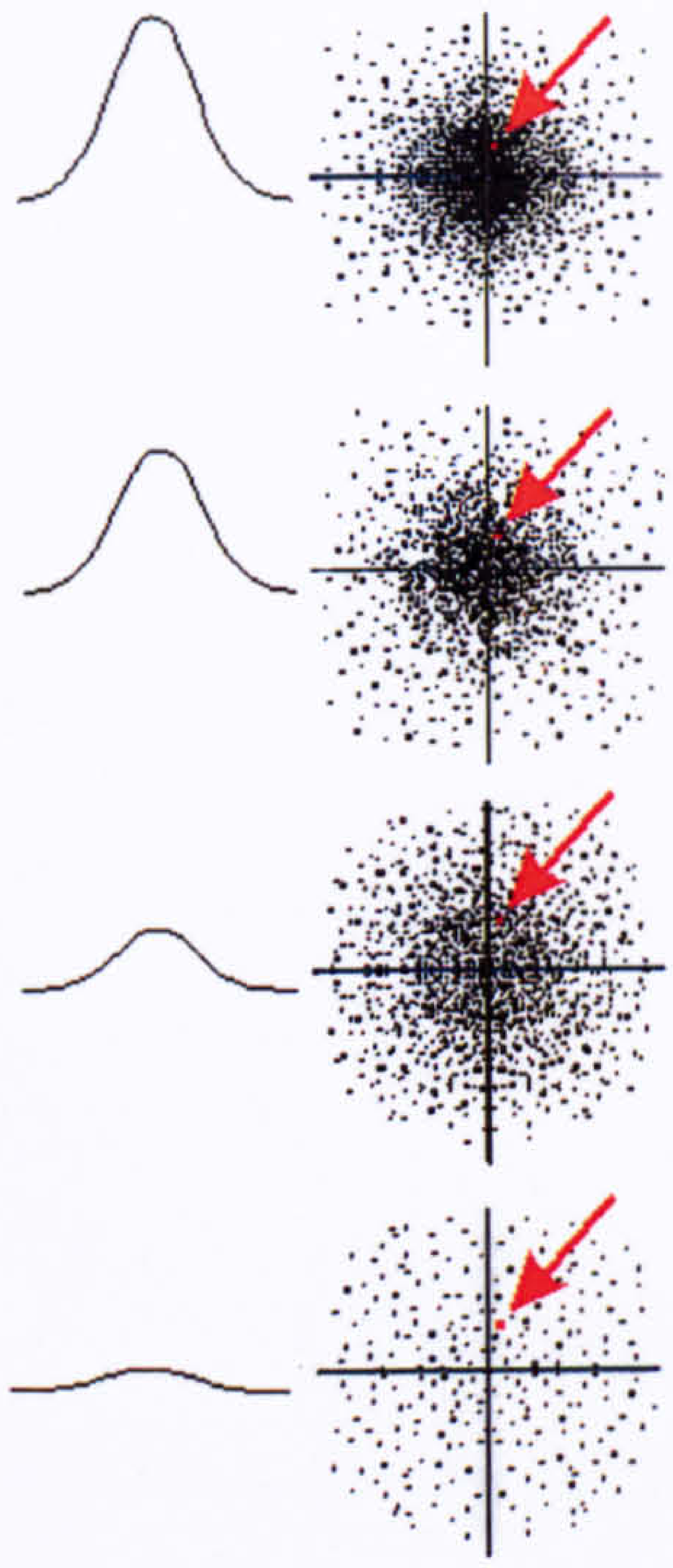


Figure 3.3. *A face once seen as distinctive (indicated by the red arrow), becomes typical as MDS becomes increasingly populated with time (right); frequency distribution reflecting distribution of representations in MDS at different points in time (left).*

By applying Sternberg's (1969) additive-factors logic to the present data, it can be inferred that the mechanisms underpinning AoA and distinctiveness are divided into serially-arranged operations or stages. Given that distinctiveness effects have been explicitly linked to FRU activity (e.g. Burton et al., 1990; Newell et al., 1999) AoA must affect familiarity decisions at a subsequent processing stage, for example at the level of PINs or the connections between FRUs and PINs. This inference is interesting given that AoA effects are increasingly seen as a ubiquitous by-product of the way in which information is stored and accessed in the brain. It also offers a refinement of Moore and Valentine's (1999) conclusion that AoA affects face processing somewhere at or before the PINs. Further, the present pattern of independent effects is most parsimoniously accommodated by a PIN-level familiarity decision mechanism, rather than one at the FRUs. This supports

Burton et al.'s (1990) contention that PINs house familiarity decisions. However, interpretation of the results to this level remains tentative. Despite having been reaffirmed since its original conception (Roberts & Sternberg, 1993; Sternberg, 1998), the general school of thought is that additive factors logic is undermined by cascade models which dispel the notion of a strict sequence.

A potential drawback of these studies, and indeed any studies which investigate the effects of AoA on face recognition, is that AoA is confounded with degree of age transformation. Unlike words, faces are dynamic, and inevitably early acquired faces will have changed in appearance to a greater degree than late acquired faces since first encountered. Whether using up-to-date images (Experiments 1 and 2) or dated images (Experiment 3), the early and late sets will be confounded with degree of age-related change. Up-to-date images of early acquired faces may not activate original representations as strongly as late acquired faces: only experiments which specifically ask whether or not AoA effects accrue as faces undergo age-related changes can determine whether this confound is problematic. Dated images of early acquired faces may be less recognisable than dated images of late acquired faces, incurring processing costs as a function of age-related change: only future studies can determine to what extent representations are ageless and to what extent recognition demands an exact match of age characteristics. One way of removing the confound between AoA and degree of age-related change, should it exist, would be to compare responses to images acquired early for one set of participants with responses to the same images presented to older participants for whom they are late acquired.

Another pertinent issue, which may have compromised results in Experiment 1 at least, is the difficulty in obtaining stimuli sufficiently familiar to the majority of participants. Particular stimuli attracted a large number of errors. For example, 71% of participants failed to identify Neil Kinnock. Indeed, Lewis (1999a) identified this as a problem inherent in face recognition studies. Compounding this problem is the idiosyncratic nature of knowledge of famous faces. For example, Ewan McGregor was not recognised by 43% of participants in Experiment 2, despite having been left in the database owing to his familiarity. Furthermore, in a later experiment (Experiment 9) Ewan McGregor was successfully recognised by 86% of participants!

Lewis (1999a) also drew attention to the problem of standardising images so that they are equally good depictions of faces. In the present series of experiments, sets of stimuli were always matched for 'likeness' so that the average level of standardisation was balanced. However, matching sets by likeness does not standardise images, it simply ensures that each set has an equal number of faces which are not such good representations of the individual and those which are better. Given that this variable dictates the extent to which recognition is impeded altogether, it does not matter how well matched the sets are on AoA, distinctiveness and frequency if the images are not good representations of the individual. Ideally, the database should have been screened to ensure that all faces were at least good likenesses. However, there were not enough faces in the database to permit such rigorous screening. To create the perfect database is certainly outside of the time constraints imposed by this research.

Despite the variation in the results of Experiments 1 to 3, probably stemming from artefacts associated with the particular stimulus sets used, there are certainly no grounds to confirm the hypothesis that AoA and distinctiveness interact. Additivity between the variables implies that the mechanisms sub-serving AoA and distinctiveness are independent and thus the current theories posited to explain these effects require no integration nor elaboration.

CHAPTER 4 – EFFECTS OF AOA AND DISTINCTIVENESS ON GENDER PROCESSING

4.1 Introduction

Interest now turns to an exploration of the relationship between AoA and gender processing, in which AoA is recruited as an investigative tool to facilitate understanding of the relationship between identity and gender processing. Specifically, Experiment 4 assesses the effect of AoA on gender classification performance with full face stimuli. This provides a prerequisite for Experiments 5 to 10, which develop this line of research by manipulating the visual stimulus as a further attempt to recover effects of AoA on gender processing. Experiment 4 replicates Experiment 2 as a gender classification task. Thus, the factorial manipulation of AoA and distinctiveness associated with this familiarity decision experiment is preserved. By virtue of this design, the effect of distinctiveness on gender processing was also explored. This subsidiary investigation does stray somewhat from the core objective of this thesis, namely to explore the cognitive operations associated with AoA. Nonetheless, it bolsters exploration of the relationship between identity and gender processing, insofar as an effect of distinctiveness opens up the possibility that MDS interfaces identity and gender processing.

The effect of AoA on gender classification performance has not yet been investigated. This is not surprising as there are no *a priori* reasons to suppose that gender classifications are made with recourse to identity. Indeed, identity and gender have historically been interpreted as independent processes, hence their functional separation in the Bruce and Young (1986) model. Moreover, Bruce (1986) and Bruce et al. (1987a) demonstrated that gender processing is not modulated by face familiarity. Bruce (1986) did, however, recover effects of familiarity on gender classifications made to two particularly androgynous faces. This is commensurate with recent studies which deliberately rendered processing conditions laborious, by removing external gender cues, and recovered effects of identity on gender processing (Clutterbuck & Johnston, 2004; Ganel & Goshen-

Gottstein, 2002; Goshen-Gottstein & Ganel, 2000; Rossion, 2002). Although these studies are subserved by a variety of mechanisms, broadly speaking they all hypothesised that when gender decisions can be made on the basis of superficial external cues present in full faces, gender decisions are not made with recourse to identity information. Conversely, when gender decisions are not based on a visual analysis of external gender cues, information from the identification route has the opportunity to influence gender processing. With this in mind, it is unlikely that AoA will affect full face gender decisions.

The effect of distinctiveness on gender classification performance has not yet been investigated either. This is somewhat surprising given that several studies have advocated the systematic representation of gender in multidimensional face-space (Johnston et al., 1997; O'Toole et al., 1998). An effect of distinctiveness on gender classifications would certainly support the contention that there is something systematic in the way that male and female faces colonise face-space. Johnston et al. and O'Toole et al. proposed that male and female representations are distilled into disparate regions of space to form gender-based clusters. Moreover, the latter study also claimed that representations are normally-distributed around each subcategory origin, rather than the androgynous central tendency forming the area of highest exemplar density.

The systematic colonisation of male and female faces in MDS opens up the possibility that MDS interfaces identity and gender processing. Baudouin and Tiberghien (2002) recruited MDS to explain the effect of gender on identity processing found in their study. When searching for a target face arbitrarily labelled with a male or female name, participants rejected faces of the opposite gender to the target more quickly than faces of the same gender. The authors hypothesised that when the perceived gender counters expectation, the cognitive system can ignore an entire region of MDS without having to probe precise places. An effect of distinctiveness on gender processing would demonstrate that MDS is capable of interfacing identity and gender processing.

Although there seems to be consensus concerning the representation of gender in MDS, O'Toole et al. (1998) and Johnston et al. (1997) make different predictions regarding the mechanisms underpinning gender decisions. Johnston et al. reasoned that gender decisions are made in a similar manner to facedness decisions (see Valentine & Bruce, 1986b). When a face is categorised according to gender it is compared to neighbouring representations which may either confirm or disconfirm decisions. As a result, typical faces should be classified according to gender faster and more accurately than distinctive faces, owing to their proximity to the subcategory origin. Conversely, O'Toole et al. deemed gender classification to depend on opposition to a contrastive category. Therefore, gender classification performance would depend on gender typicality (masculinity or femininity) rather than distinctiveness per se. Following this logic, some distinctive faces should be classified according to gender faster than typical faces (those furthest from the contrastive category) and some distinctive faces should be classified more slowly (those nearest the androgynous central tendency of the dimension). Thus, according to O'Toole et al., distinctiveness should not affect gender classifications. As these different mechanisms make different predictions regarding the impact of distinctiveness on gender decisions, an effect of distinctiveness in the present experiment would not only open up the possibility that MDS interfaces identity and gender processing, but it would also constrain interpretations of how gender is coded in MDS.

In Experiment 4, gender decisions were made to sets of early acquired distinctive faces, early acquired typical faces, late acquired distinctive faces and late acquired typical faces. Given that the literature boasts an independence of identity and gender processing when gender is readily discernible from physical cues such as hair style, AoA was not expected to influence full face gender classification performance. Additionally, it was predicted that typical faces would be classified according to gender faster than distinctive faces. An effect of distinctiveness on gender classification performance would open up the possibility that MDS interfaces identity and gender processing and concomitantly offer support for the mechanisms underpinning gender decisions hypothesised by Johnston et al. (1997).

4.2 Method

Participants. Twenty five undergraduate students from the University of York (17 female, 8 male) participated in this experiment, in return for course credit or a payment of £2. All participants had lived in the UK for at least 18 years and the age range tested was 18 – 21 years old. The mean age of participants was 19.40 years old (SD = 1.00).

Design. This 2 x 2 within-subjects design constituted a factorial manipulation of AoA (with two levels, early and late) and distinctiveness (with two levels, distinctive and typical). This led to the formation of four sets: early acquired distinctive, early acquired typical, late acquired distinctive, late acquired typical. A gender classification task yielded two dependent variables; response latency and response accuracy.

Materials. The stimulus set comprised the 72 famous faces featured in Experiment 2 (18 early acquired distinctive, 18 early acquired typical, 18 late acquired distinctive, 18 late acquired typical). See page 106 for details concerning the matching of these experimental sets. To preclude gender bias effects, 6 filler items were added to each set in order to equate the number of male and female faces. This increased the set size to 24 (12 males and 12 females). The analysis is based on data from the original 72 items (40 males and 32 females).

Famous faces were matched with an equal number of unfamiliar faces using gender, race and approximate age as controls. The unfamiliar faces received the same digital treatment as the famous faces.

Procedure and Apparatus. The gender classification task was run on PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is male or female" in response to the images, which were presented in the centre of the screen. Each experimental trial began with a 500 ms delay prior to image onset. Faces were displayed until a response was made. Participants logged 'male' and 'female' responses by pressing the 'z' and '/' keys on the keyboard with index fingers and these response were counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 pixels and their dimensions subtended approximately 70 x 150 mm. Participants viewed these from a distance of approximately 30cm.

Stimuli were presented in four discrete blocks, however only two were discernible. This subdivision of blocks minimized the likelihood of associative priming[19]. Stimuli were randomized within blocks and blocks were rotated between participants. During the break, participants were reminded to "respond as quickly and accurately as possible".

The experiment commenced with 40 practice trials (20 male and 20 female) and presentation of the experimental trials was directly preceded by eight lead-in trials. Both practice and lead-in trials contained a representative sample of stimuli (early and late acquired, typical and distinctive, famous and unfamiliar, male and female) to preclude the possibility of bias effects. None of these items were repeated as experimental trials or analysed. The task took approximately 10 minutes.

4.3 Results

All participants classified at least 75% of the famous faces correctly and reached an accuracy threshold of 75% correct when famous and unfamiliar responses were pooled together. Thus, no participants were replaced.

The accuracy analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications; b) reaction times deemed outliers. As a result of incorrect classifications of famous faces, 51 (2.8%) of the responses were removed. Forty nine (2.7%) outliers were removed

19 However, associative priming is an unlikely product of a sex classification task. Ganel and Goshen-Gottstein (2000) found that repetition priming occurs under strict conditions: The prime must be a whole face or internal features; internal features must be presented at test; the study task must guide attention towards internal features e.g. intelligence rating task. As no emphasis was placed on internal features it is *a priori* likely that associative priming would not occur.

for being slower than 3 standard deviations away from the mean or faster than 300ms.

The latency analysis was carried out on correct responses only. Table 4.1 shows the mean correct response latencies in each condition of the experiment and percentage error rates. Unfamiliar face summary data is also reported. For inspection purposes, '% misclassifications', referring to the percentage of incorrect gender classifications, are documented.

Table 4.1. *Mean RTs and Percentage Errors for Gender Decisions in Each Condition (Experiment 4).*

	Early Acquired		Late Acquired		Unfamiliar
	Distinctive	Typical	Distinctive	Typical	
Mean RT	551	535	552	524	536
SD	(86)	(86)	(93)	(82)	(86)
% Error	6.67	6.00	4.67	4.89	4.78
% Misclass.	3.78	2.89	2.22	2.44	2.22

Reaction Times. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. For this and subsequent analyses, by-subject data was submitted to post-hoc tests and reported means stem from this analysis. The main effect of AoA was not significant in the by-subject analysis, $F_1(1,24) = 1.33$, MSe = 554.42, $p = .26$, or the by-items analysis, $F_2(1,68) = 0.40$, MSe = 1,719.75, $p = .53$. However, the main effect of distinctiveness reached significance in both analyses, $F_1(1,24) = 16.17$, MSe = 756.44, $p<.001$; $F_2(1,68) = 6.09$, MSe = 1,719.75, $p<.05$, confirming the observed tendency for typical faces (529 ms) to be classified according to gender faster than distinctive faces (551 ms). The interaction between AoA and distinctiveness was not significant in either analysis, $F_1(1,24) = 0.98$, MSe = 846.90, p = .33; $F_2(1,68) = 0.27$, MSe = 1,719.55, p = .60.

A post-hoc analysis checked to see whether or not the effect of distinctiveness was confounded with gender typicality. It is possible that typical faces are classified according to gender more quickly than distinctive ones because they are more gender typical. Taking advantage of gender typicality ratings collected for later experiments (see p. 210), independent t-tests investigated how well the sets were matched for this variable. The sets were matched (in each case $t \leq -1.43$), with exception to a significant difference between the early and late typical sets, $t(34) = -2.33$, $p<.05$. Importantly, the gender typicality of faces in the distinctive sets and the typical sets was not significantly different. Therefore, the above results appeared not to be an artefact of gender typicality. Further, an analysis of covariance, carried out with AoA and distinctiveness as between-subjects factors and gender typicality as a covariate, demonstrated that although gender typicality influenced gender classification latency, as one would expect, $F_2(1,67) = 19.88$, MSe = 1,346.11, $p<.001$, distinctiveness affected performance in its own right, $F_2(1,67) = 8.35$, MSe = 1,346.11, $p<.01$. To compare the magnitude of the distinctiveness effect before and after gender typicality was entered into the analysis as a covariate, distinctiveness effect sizes are reported. Indexed by Pearson's correlation coefficient, the effect size was similar before covariance ($r = .30$), and after ($r = .33$). In both instances, distinctiveness could be termed a "medium" sized effect (Cohen, 1988).

Errors. The data were analysed by-subjects in a repeated measures ANOVA with AoA (early or late) and distinctiveness (distinctive or typical) as within-subjects factors, and by-items with AoA and distinctiveness as between-subjects factors. The main effect of AoA was not significant by-subjects or by-items, $F_1(1,24) = 1.45$, MSe = 1.36, $p = .24$; $F_2(1,68) = 1.09$, MSe = 2.50, $p = .30$. The main effect of distinctiveness also failed to attain significance, $F_1(1,24) = 0.07$, MSe = 0.56, $p = .79$; $F_2(1,68) = 0.02$, MSe = 2.50, $p = .88$. The interaction between AoA and distinctiveness was not significant in either analysis, $F_1(1,24) = 0.25$, MSe = 0.64, $p = .62$; $F_2(1,68) = 0.09$, MSe = 2.50, $p = .77$.

4.3.1 Participant Gender and Face Gender Analysis

To assess whether effects in the latency analysis varied as a function of participant gender, this factor was inserted into the analysis as a between-subjects factor. Only those results which indicate some systematic effect on the experimental variables are reported. Although the main effect itself was not significant, participant gender interacted significantly with distinctiveness, $F_1(1,23) = 5.45$, MSe = 638.14, $p<.05$, with males tending to show a larger distinctiveness effect (39 ms) compared with females (14 ms). However, a simple main effects analysis (used instead of Tukey tests because male and female group sizes varied) found no interactions. The effect of face gender was also inserted into the analysis as a within-subjects factor but failed modulate the latency analysis. Thus, effects in the latency analysis appeared not to vary markedly as a function of participant gender or face gender.

4.3.2 Familiarity Analysis

Post-hoc analyses were carried out on the famous face and unfamiliar face data to provide an additional means of assessing the impact of familiarity on gender decisions.

For the unfamiliar faces, the accuracy analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications; b) reaction times deemed outliers. As a result of incorrect classifications, 40 (2.2%) of the responses were removed. Forty six (2.6%) outliers were removed for being slower than three standard deviations away from the mean or faster than 300ms.

Only correct gender decisions were submitted to the latency analysis. See Table 4.1 for mean famous and unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects in a repeated measures ANOVA with familiarity (famous or unfamiliar) as within-subjects factors, and by-items with familiarity as a between-subjects factor. The effect was not significant in either analysis, $F_1(1,24) = 1.34$, MSe = 200.33, $p = .26$; $F_2(1,142) = 0.73$, MSe = 1,379.99, $p = .39$. That is, there was no indication that gender decisions were faster to famous faces than to unfamiliar faces, which would presumably be expected if reaction times were to be modulated by AoA.

Errors. The data were analysed by-subjects in a repeated measures ANOVA with familiarity (famous or unfamiliar) as within-subjects factors, and by-items with familiarity as a between-subjects factor. Reflecting the latency analysis, the effect was not significant by-subjects, $F_1(1,24) = 1.77$, MSe = 2.21, $p = .20$, or by-items, $F_2(1,142) = 0.79$, MSe = 1.72, $p = .38$.

4.4 General Discussion

Full-face gender classification latencies were not affected by AoA. However as predicted, distinctiveness affected performance such that typical faces were classified faster than distinctive faces. No effects were found in the accuracy analysis, perhaps because error rates were so low: averaged across conditions, 5.6% famous faces were misclassified compared with an error rate of 13.0% to the same faces classified according to familiarity in Experiment 2.

The absence of AoA effects on full face gender classification is compatible with the functional separation of identity and gender processing in Bruce and Young's (1986) model. The notion that the cognitive operations are independent is also supported by the post-hoc finding that gender classification performance is not affected by face familiarity. Furthermore, these findings are not surprising in light of research which reports independent effects of identity and gender processing when gender decisions can be made from a visual analysis of sexually dimorphic cues (Bruce, 1986; Bruce et al., 1987a; Ganel & Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000). However, these studies also hypothesised that the identification route can influence gender decisions when gender can no longer be determined from a visual inspection of gender cues. This may arise specifically when external cues are removed (Ganel & Goshen-Gottstein, 2002; Goshen-

Gottstein & Ganel, 2000) or simply if a face is intrinsically androgynous (Bruce, 1986; Clutterbuck & Johnston, 2004; Rossion, 2002). Experiments 5 to 10 continue to investigate the relationship between identity and gender processing by manipulating the visual stimulus to remove external cues to gender or vary gender typicality.

Interpreted with regard to MDS, the effect of distinctiveness on gender processing opens the possibility that MDS interfaces identity and gender processing: there is clearly something systematic in the way that male and female faces colonise face-space. This is consistent with proposals made by Johnston et al. (1997) and O'Toole et al. (1998) which state that MDS is characterised by two gender-based clusters. Moreover, this notion also sits comfortably with Baudouin and Tiberghien's (2002) hypothesis that knowing the gender of a face facilitates identification. As the authors suggested, it is plausible that the cognitive system screens off regions of face-space to restrict the search for a target face to a region of space corresponding to the appropriate gender. This subsidiary investigation certainly offers one possible locus where identity and gender processing are not entirely independent.

The effect of distinctiveness on gender decisions also constrains interpretations of the mechanisms posited to underpin gender decisions. This result is consistent with Johnston et al.'s (1997) hypothesis that gender classifications are made by comparing the stimulus to neighbouring representations which either confirm or disconfirm the decision. This hypothesis predicts that typical faces should be classified according to gender faster than distinctive faces by virtue of their close proximity to other exemplars of the same gender. Alternative face-space theories which hypothesise that a prototype or norm is abstracted, make the same prediction as exemplar-based models. Whereas exemplar-based models appeal to exemplar density to explain the effect of distinctiveness on gender decisions, norm-based models appeal to distance from the norm. In the latter type of model, typical faces are classified according to gender faster than distinctive faces because they are most similar to the prototypical male or female face. Although the present findings cannot distinguish between exemplar- and norm-based conceptualisations, they do indicate that the mechanisms hypothesised by O'Toole et al. (1998) are unlikely.

O'Toole et al. suggested that classification depends on degree of opposition to a contrastive category; a measure most aptly captured by gender typicality. However following this logic, distinctiveness effects should be absent or at least attenuated. Furthermore, when gender typicality was entered as a covariate in the latency analysis, effects of distinctiveness persisted in addition to those attributed to gender typicality.

Interpreted with regard to the perceptual representations extracted during identity and gender processing, the customary effect of distinctiveness on familiarity decisions (Experiments 1 to 3) and reversed effect of distinctiveness (Experiment 4) is a conceivable pattern. If identity and gender processing were based on identical perceptual representations, then those same cues which make a typical face easy to classify according to gender need also be responsible for making that face difficult to recognise. That faces do not appear to be represented by identical representations, means that cues do not need to assume this complex role of dual responsibility: whilst eyes and brows appear salient for identity processing (O'Donnell & Bruce, 2001; Roberts & Bruce, 1988; Sadr et al., 2003) and gender processing (Brown & Perrett, 1993; Bruce et al., 1993a; Burton et al., 1993), 3D structure and facial protuberance of particular features, has been found particularly important for gender processing (Bruce et al., 1993a; Burton et al., 1993; Bruce & Langton, 1994; Enlow, 1982).

As a potential drawback, it is possible that the emergence of AoA effects is impeded by the fact that the latency analysis incorporated responses to faces unknown to certain participants. In a familiarity decision task, the latency analysis is based on responses to known faces because a face is usually misclassified if it unknown. However, in a gender classification task, misclassifications do not reveal anything about whether or not a face is known and so the latency analysis may include responses to unfamiliar faces. In a paradigm designed to explore the relationship between identity and gender processing, it is vital that the latency analysis is based on responses to known faces. In response, the forthcoming experiments implement familiarity verification sessions as a means of indicating which faces are known.

The classification of full faces according to gender was not affected by AoA. Therefore, the relationship between identity and gender processing appears to be one of independence, at least for full faces. As predicted, a reversed distinctiveness effect characterised gender classification performance. This is consistent with the systematic accommodation of male and female faces in MDS and indicates that MDS may provide one locus for the integrated processing of identity and gender.

CHAPTER 5 – THE RELATIONSHIP BETWEEN IDENTITY AND GENDER PROCESSING: RECRUITING AOA AS AN INVESTIGATIVE TOOL

5.1 Introduction: Experiments 5 to 8

The cognitive operations associated with identification and gender processing were cast as functionally separate in Bruce and Young's (1986) model of face recognition. Whereas the model's heterarchic structure is largely supported by double dissociations between neuropsychological case studies, the functional separability of identity and gender processing is only underpinned by a single dissociation (Tranel et al., 1988). Inevitably, the controversial status of this relationship is often cited as a trigger for research (e.g. Ganel & Goshen-Gottstein, 2002; Rossion, 2002). The purpose of Experiments 5 to 8 was to ascertain the nature of the relationship between identity and gender processing. These experiments recruited AoA as an investigative tool to index effects of familiarity on gender processing. As a by-product of this investigation, it was also possible to measure AoA effects using gender classification tasks which do not require overt recognition of the target face and owing to their simplicity, are less vulnerable to errors. Experiments 5 to 8 did not seek to explore the various mechanisms hypothesised to underpin this relationship. Rather, this is addressed in Chapter 7. Nevertheless, parallel-route and single-route accounts are outlined briefly in order to inform predictions.

Studies advocating parallel-route accounts either found independent effects of identity and gender processing (Bruce et al., 1987a), or an effect of identity on gender processing which arose when gender processing was laborious (Bruce, 1986; McNeil et al., 2003). Bruce attributed an effect of familiarity on gender processing to two particularly androgynous items and concluded tentatively that when the analysis of sexually dimorphic features is slowed, the cognitive system accesses semantics to facilitate the decision. In addition, McNeil et al. found that the absence or presence of repetition priming onto gender decisions depended on the locus in the system where the decision was made. Gender decisions to full faces and full names were not susceptible to priming because superficial gender cues enabled judgments to be made without recourse to identity information. However, surnames were primed, leading McNeil et al. to conclude that the semantic system can be recruited to facilitate gender decisions. Further, Bruce et al. observed that gender decisions are usually made sufficiently quickly to negate semantic support and so the time courses of identity and gender processing must be in the same range to support an interaction between operations.

Studies by Rossion (2002) and Clutterbuck and Johnston (2004) reported robust effects of familiarity on gender processing for nearly all items, with significant effects of familiarity in by-subjects and by-items analyses. To explain the differences between these results and Bruce's (1986) less robust finding, Rossion suggested that whilst his stimuli were completely cropped to remove superficial cues to gender, cues in the earlier study were minimal but still available for gender processing. Nevertheless, despite rendering gender processing laborious, response times in Rossion's study (approximately 695 ms) and those reported by Clutterbuck and Johnston (approximately 595 ms) were fast. According to Clutterbuck and Johnston "the fact that reaction times are still very fast makes semantic access unlikely" (p. 165). Both studies hypothesised modifications to the original parallel-route account to explain effects of identity on gender processing.

Rossion (2002) found that morphs of familiar and unfamiliar faces with a 60-100% familiar face contribution were classified according to gender faster than morphs which were saturated with a 0-40% contribution. He concluded that the cognitive system must have discriminated familiar and unfamiliar faces prior to the gender decision because the effect was categorical and not dependent on the degree of visual similarity between the familiar face and the morph. To accommodate a relationship between these operations where semantic activation is initiated before earlier processes are completed, Rossion hypothesized a cascade implementation of the parallel-route model. As an alternative explanation for this interdependence of processes, Clutterbuck and Johnston (2004) offered a modification of Burton et al.'s (1990) IAC model of face recognition. Made possible by the bi-directional links fundamental to IAC architecture, they hypothesised that FRUs pass feedback to structural encoding units to stabilize the representations they contain. Therefore, familiar faces receive FRU feedback plus information from structural encoding to facilitate gender processing, compared with unfamiliar faces which rely solely on the analysis of facial structure. Both of these hypotheses preserve the characteristic flavour of parallel-route models whilst bypassing semantic access problems.

Goshen-Gottstein and Ganel (2000) and Ganel and Goshen-Gottstein (2002) pioneered the single-route hypothesis which predicts that identity and gender are processed together, and entertains a false separability between operations when superficial 'hair-style heuristics' dictate gender decisions. The earlier study demonstrated that gender decisions made to familiar and unfamiliar hair-deleted faces were susceptible to repetition priming. The later study found that the dimensions of identity and gender could not be selectively attended to in a Garner-Interference task with hair-deleted faces. Together, these studies argued that the dimensions are processed together at perceptual and mnemonic stages in processing. FRUs were hypothesised to mediate identity and gender processing but the authors hypothesised that engagement of FRUs rested critically on focused attention to internal features. Superficial hair-style heuristics were hypothesised to truncate the perceptual whole and negate FRU engagement. Whereas studies in favour of parallel-route accounts (Bruce, 1986; Clutterbuck & Johnston, 2004; McNeil et al., 2003; Rossion, 2002) argued that identity affects gender processing under suboptimal conditions, Ganel and Goshen-Gottstein proposed that an integrality of cognitive operations is the default and false separability emerges when gender is judged on the basis of superficial cues.

Although radical, a single-route hypothesis is possible given that there is some overlap in the perceptual representations which mediate identification (e.g. O'Donnell & Bruce, 2001; Sadr et al., 2003) and support gender decisions (e.g. Brown & Perrett, 1993; Burton et al., 1993). Moreover, the idea of overlapping representations is compatible with Haxby et al.'s (2000) proposition that the fusiform face area processes invariant physiognomic information, which includes identity and gender.

The aim of Experiments 5 to 8 was to ascertain the nature of the relationship between identity and gender processing, and the conditions under which identity affects gender processing. Gender decisions were made to early acquired and late acquired full faces, internal features (eyes, nose, mouth) and external features (hairstyle, jawline, chin). To ensure that the stimulus sets employed readily yielded AoA effects, familiarity decision tasks (Experiments 5 and 7) were administered prior to the gender classification tasks (Experiments 6 and 8). Each of the experiments included a familiarity verification session during which participants indicated which faces were actually known to them. Likewise, Clutterbuck and Johnston (2004) administered a familiarity verification session after their gender classification task to ensure that the latency analysis was carried out solely on responses to recognised faces.

AoA was hypothesised to affect familiarity decisions made to whole faces, internal features and external features, with early acquired faces classified more quickly and accurately than late acquired faces. There was no reason to suspect that this manipulation of the visual image would modulate AoA effects because, in Lewis' (1999) terms, AoA is a nonpictorial factor: the effect is not rooted in properties of the visual image and should not be affected by manipulations at this level of processing. Conversely, it was hypothesised that AoA effects in the gender classification experiment would be restricted to the internal features condition. Superficial cues to gender visible in whole faces and external features can facilitate gender decisions without recourse to identity information.

To provide further evidence that familiarity influences gender processing, a post-hoc familiarity analysis compared gender classifications made to famous and unfamiliar faces. It was hypothesised that in the internal feature condition, gender classifications made to famous faces would be facilitated relative to unfamiliar face performance. Similarly, Goshen-Gottstein and Ganel (2000) carried out a post-hoc familiarity analysis to "help determine the validity of the parallel-route hypothesis" (p. 1207). Whilst they found that familiarity with a face facilitated gender decisions, this was true for both full faces and hair-deleted faces. Given their suggestion that a false separability between identity and gender processing emerges when superficial hair-style heuristics abound, it is surprising that full faces supported familiarity effects. Further, it is surprising that Goshen-Gottstein and

Ganel do not acknowledge this finding. Nevertheless, familiarity is hypothesised to affect gender decisions made to internal features in the forthcoming experiments.

5.2 Experiment 5

5.2.1 Introduction

Experiment 5 investigated the effect of AoA on familiarity decisions to ensure that the stimulus set readily yielded AoA effects. Familiarity decisions were made to early acquired and late acquired faces taken from the current database. These were presented once as whole faces, once as internal features and once as external features. A familiarity verification session administered after the experiment proper ensured that the latency analysis was based on responses only to faces genuinely recognised by participants.

AoA was hypothesised to affect face processing regardless of face type. Commensurate with this, it was hypothesised that famous faces would be classified according to familiarity faster than unfamiliar faces in the post-hoc familiarity analysis: Bruce et al. (1987a) found that familiarity modulated performance in this manner. This analysis was carried out to facilitate comparisons with the post-hoc familiarity analysis of gender classifications. However, the value of this task is limited because it is suboptimal to compare responses to categories which correspond to separate responses in a binary decision task (Brysbaert et al., 2000). Also, the accuracy analysis is limited because errors reflect response strategies in addition to intrinsic properties of the target faces: the tendency to respond 'unfamiliar' when uncertain can only improve accuracy with unfamiliar faces and adversely affect famous face performance.

5.2.2 Method

Participants. Thirty six undergraduate students from the University of York took part in this experiment (8 male, 28 female), receiving course credit or a payment of £2 for their participation. Participants were 18 – 22 years old (mean = 19.58; S.D. = 0.94) and were required to have lived in the UK for at least 18 years.

Design. The experimental variables comprised face type (with three levels; whole face, internal features or external features), AoA (with two levels; early or late) and gender of face (with two levels; male or female) which were explored in a repeated measures design. The order of the face type conditions was counterbalanced across participants yielding six different predefined orders of presentation.

The dependent variables, response latency and accuracy, were the products of a familiarity decision task.

Materials. Sixty eight faces (34 early acquired, 34 late acquired) were selected from the current database. At each level of AoA, there was an equal number of male and female faces. Independent t-tests confirmed that the four sets (early acquired male, early acquired female, late acquired male and late acquired female) were matched for frequency, distinctiveness, likeness and familiarity, and the two early sets and two late sets were each matched for AoA (in each case $t \leq .898$). See Table 5.1 for summary data pertaining to the creation of experimental sets.

Table 5.1. *Mean Ratings as a Function of Experimental Set (Experiment 5).*

	Early acq. Male			Late acq. Male			Early acq. Female			Late acq. & female		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.85	(0.41)	2.14 - 3.41	4.06	(0.49)	3.57 - 5.00	2.76	(0.48)	1.74 - 3.43	4.07	(0.47)	3.48 - 5.00
Distinctiveness	2.92	(0.92)	1.65 - 4.80	2.97	(0.74)	1.60 - 4.13	2.96	(0.57)	1.80 - 3.80	2.91	(0.59)	2.05 - 4.35
Frequency	2.53	(0.46)	2.00 - 3.65	2.62	(0.52)	1.85 - 3.70	2.72	(0.75)	1.70 - 3.95	2.68	(0.53)	1.43 - 3.40
Likeness	3.97	(0.34)	3.00 - 4.62	4.00	(0.37)	3.85 - 4.71	4.03	(0.39)	3.30 - 4.62	4.04	(0.32)	3.25 - 4.52
Familiarity	4.40	(0.41)	3.68 - 4.95	4.42	(0.24)	3.95 - 4.82	4.39	(0.41)	3.73 - 4.95	4.40	(0.31)	3.95 - 4.82

Using the Adobe PhotoShop 7.0 package, faces in the internal condition were created by deleting hair and contours from the whole face images, so that only the internal features (eyes, nose and mouth) remained. All internal images were equated in size and were captured in an oval template subtending 180 x 210 pixels. External images were the complement of internal images, comprising hair, ears and face-shape. Examples of the stimuli seen in each of the three conditions are shown in Figure 5.1.

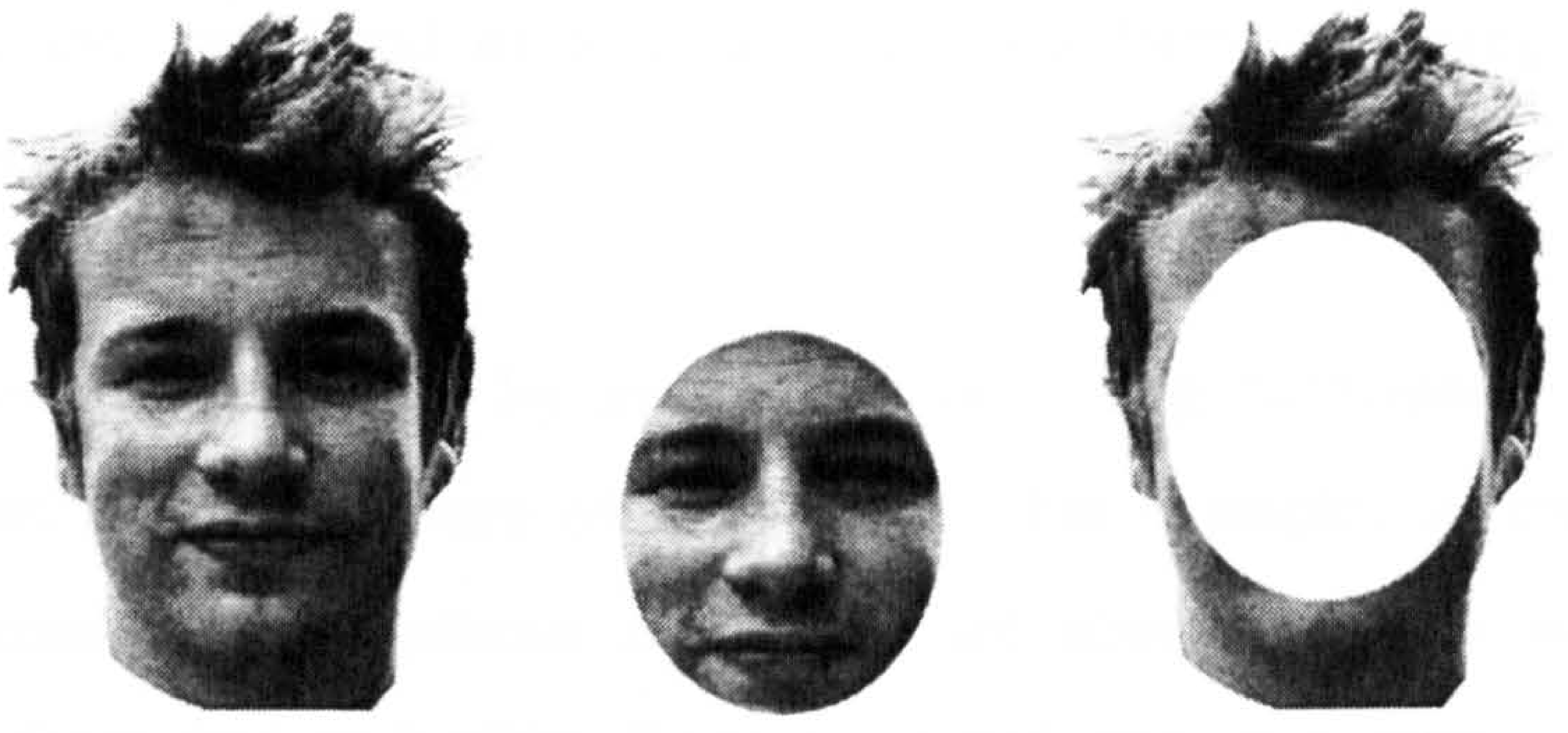


Figure 5.1. *Examples of the stimuli used in the whole (left), internal (middle) and external (right) conditions of Experiments 5 to 8: comparable to stimuli in previous experiments (e.g. Young et al., 1985; A. Ellis, Burton, Young & Flude, 1997).*

The experimental stimuli were matched with an equal number of unfamiliar faces on gender, approximate age and racial group. These images were prepared in the same manner as the famous faces.

Procedure and Apparatus. The familiarity decision task was executed in PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to the images presented in the centre of the screen. Each experimental trial began with a 500 ms blank screen prior to image onset. Faces were displayed until a response was made. 'Familiar' and 'unfamiliar' responses were logged using the 'z' and '/' keys on the keyboard, and keys pressed with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 pixels and the absolute dimensions of the whole face images were approximately 70 x 150 mm. Participants viewed these from a distance of approximately 30cm.

The 136 experimental trials (68 famous and 68 unfamiliar) were presented once as whole faces, once as internal features and once as external features. The face type conditions were blocked, and the order of the blocks was counterbalanced yielding six orders of presentation: whole, internal, external (WIE); whole, external, internal (WEI); internal, whole, external (IWE); internal, external, whole (IEW); external, whole, internal (EWI); external, internal, whole (EIW). Each face type block was further subdivided into two blocks and the order of the two sub-blocks was randomized between participants. This subdivision was introduced into the design to reduce the likelihood of associated famous faces appearing in succession, causing associative priming.

Each block was preceded by instructions informing participants about the task requirements and the nature of the stimuli. For example, an excerpt from the internal condition instructions read "you are about to see a series of PART celebrity faces and unfamiliar faces presented one at a time. The parts will comprise internal features (eyes, nose, mouth). See example below". The instructions preceding the second and third blocks varied slightly from those read at the start as participants were additionally informed that the familiarity decision referred specifically to whether or not the face was famous. This line was inserted to counteract a growing sense of familiarity for the unfamiliar faces stemming from their repeated presentation over three blocks.

Forty practice items (20 famous and 20 unfamiliar) were administered prior to each of the three blocks. After the practice session, participants were reminded to respond "as quickly and accurately as possible". The experimental trials were directly preceded by 8 lead-in trials as a means of absorbing any initial hesitations. A representative sample of stimuli (early and late acquired; male and female faces) constituted practice and lead-in items to eliminate the possibility of bias effects. None of these items were repeated as experimental items or included in the subsequent analysis. Practice and lead-in images were presented as whole faces, internal features or external features as appropriate.

After the experiment, participants completed a familiarity verification exercise. Participants were presented with the intact versions of the 136 faces and asked to indicate at their own pace "which faces are familiar to you (i.e. you recognize them as being famous)". They were explicitly informed that the task was not timed and accuracy was paramount. Coupled with the experiment proper, the session lasted approximately 25 minutes.

5.2.3 Results

Seven participants had to be replaced for failing to meet the criteria which required participants to: a) produce 75% correct responses to famous faces in the whole face condition and familiarity verification session (to ensure that the latency analysis is based on sufficient responses); b) produce 75% correct responses to famous faces and unfamiliar faces pooled together, in the whole face condition and familiarity verification session (to prevent inclusion of data derived from response strategies); c) produce 75% correct responses to unfamiliar faces in the internal and external conditions (to ensure that despite the high error rates to famous faces in these conditions, correct 'familiar' responses are likely to be valid and not the product of a response strategy).

The error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions made to known faces (a face is 'known' if it is successfully recognized in the familiarity verification session); b) a failure to recognise a face in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect familiarity decisions made to known faces, 158 (6.5%) whole face responses, 607 (24.8%) internal feature responses, and 739 (30.2%) responses made to external features, were removed. In addition, 111 (4.5%) responses from each face type were removed as a consequence of failing to recognise faces in the familiarity verification session. As a result of eliminating outliers slower than 3 standard deviations away from the mean and faster than 300ms, 44 (1.8%) whole face responses were removed, as were 34 (1.4%) from the internal condition and 33 (1.4%) from the external condition. Outliers were calculated for each face type separately.

Only correct familiarity decisions and those made to faces successfully recognized in the familiarity verification session, were submitted to the latency analysis. Table 5.2 shows the mean correct response latencies in each condition of the experiment in addition to associated error rates. For inspection purposes, Table 5.2 additionally reports '% misclassifications' which denotes incorrect familiarity decisions made to known faces.

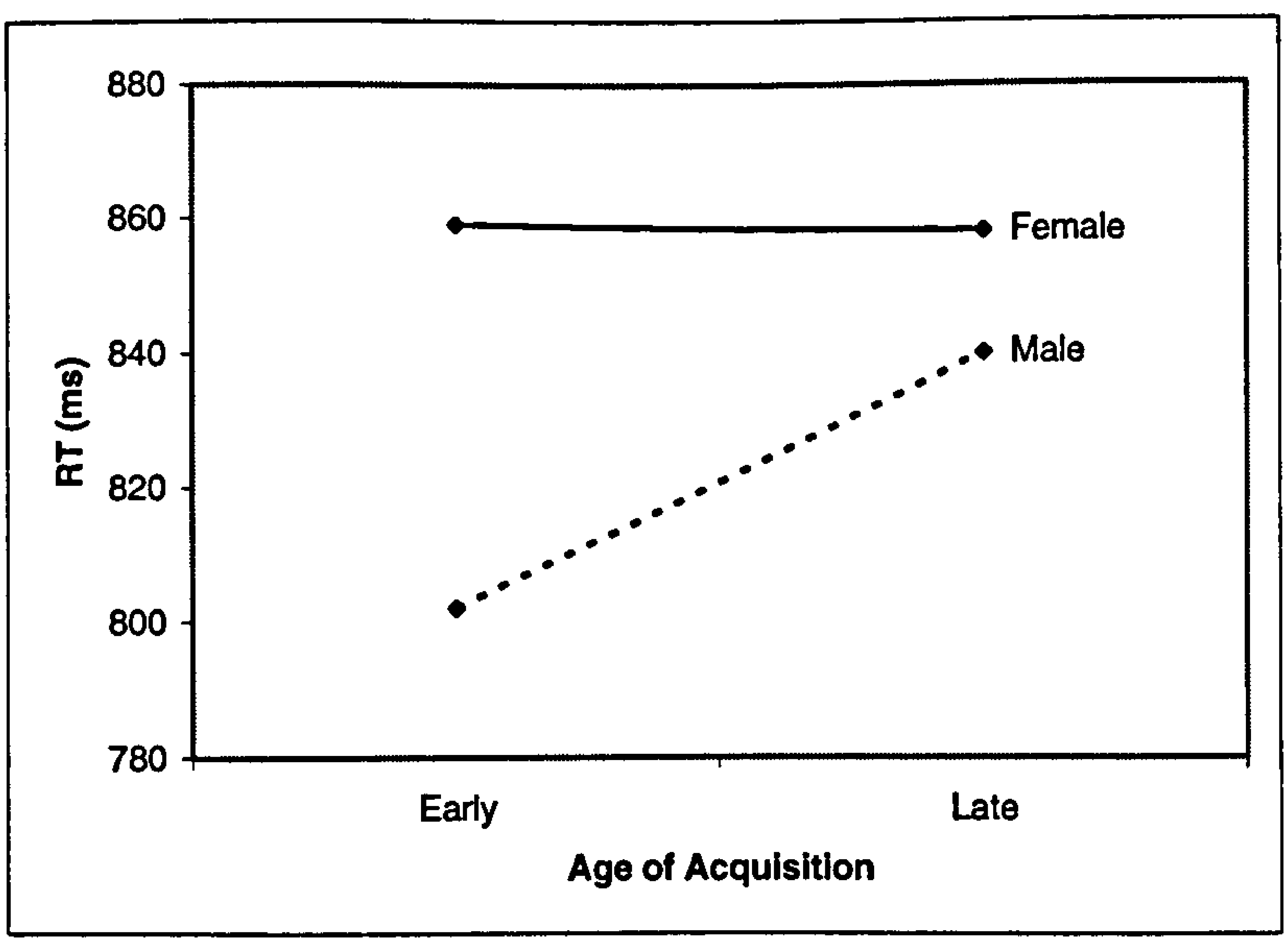
Table 5.2. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Experiment 5).*

		Early Acquired			Late acquired			Unfamiliar		
Stimulus		Male	Female	Mean	Male	Female	Mean	Male	Female	Mean
Whole	*Mean RT*	672	731	700	702	714	707	702	737	719
	SD	101	140	114	99	143	116	123	127	123
	% Error	10.5	14.9	12.7	15.7	10.1	12.9	6.2	6.7	6.5
	% Misclass.	5.7	8.2	7.0	8.7	4.9	6.7	1.9	1.9	1.9
Internal	*Mean RT*	829	885	857	869	906	885	856	896	876
	SD	146	164	150	165	221	184	171	194	180
	% Error	29.2	30.7	30.1	33.2	29.6	31.4	8.9	10.3	9.6
	% Misclass.	25.4	26.2	25.9	28.2	25.1	26.7	4.8	5.6	5.2
External	*Mean RT*	903	960	927	949	955	951	896	918	907
	SD	186	237	197	230	207	216	182	190	181
	% Error	25.3	47.1	36.2	37.7	34.5	36.1	12.4	13.8	13.1
	% Misclass.	21.3	43.4	32.2	32.5	30.3	31.4	8.5	9.6	9.0

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors, and by-items with face type, AoA and face gender as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of face type was significant in both analyses, $F_1(2,70) = 41.84$, MSe = 51,191.95, $p<.001$; $F_2(2,192) = 124.41$, MSe = 8,663.58, $p<.001$, confirming the observed tendency for wholes (705 ms) to be recognized more quickly than internals (872 ms), and internals more quickly than externals (942 ms). With alpha set at .05, Tukey tests revealed that whole faces were recognised significantly faster than internal and external features, but the difference between the latter face types was not significant (HSD = 128). The main effect of AoA was significant by-subjects, $F_1(1,35) = 8.48$, MSe = 4,681.80, $p<.01$, but not by-items, $F_2(1,192) = .49$, MSe = 8,663.58, $p = .49$, with early acquired faces (830 ms) recognized faster than late acquired faces (849 ms). The main effect of gender was significant by-subjects, $F_1(1, 35) = 16.55$, MSe = 9,302.90, $p<.001$, and approached significance by-items, $F_2(1,192) = 3.44$, MSe = 8,663.58, $p = .065$, stemming from faster familiarity decision responses to male faces (821 ms) compared with females (858 ms).

The two-way interaction between AoA and gender was significant in both analyses, $F_1(1,35) = 9.47$, MSe = 4,222.89, $p<.01$; $F_2(1,192) = 6.70$, MSe = 8,663.58, $p = .01$, see Figure 5.2. *A posteriori* Tukey's comparisons (HSD = 41) demonstrated that AoA did not affect the recognition of females, whereas the effect approached significance for male faces. Whilst there was no statistically significant effect of gender on late acquired faces, a significant difference was found between the speed at which early male and female faces were recognized. Neither analysis yielded any further significant interactions.

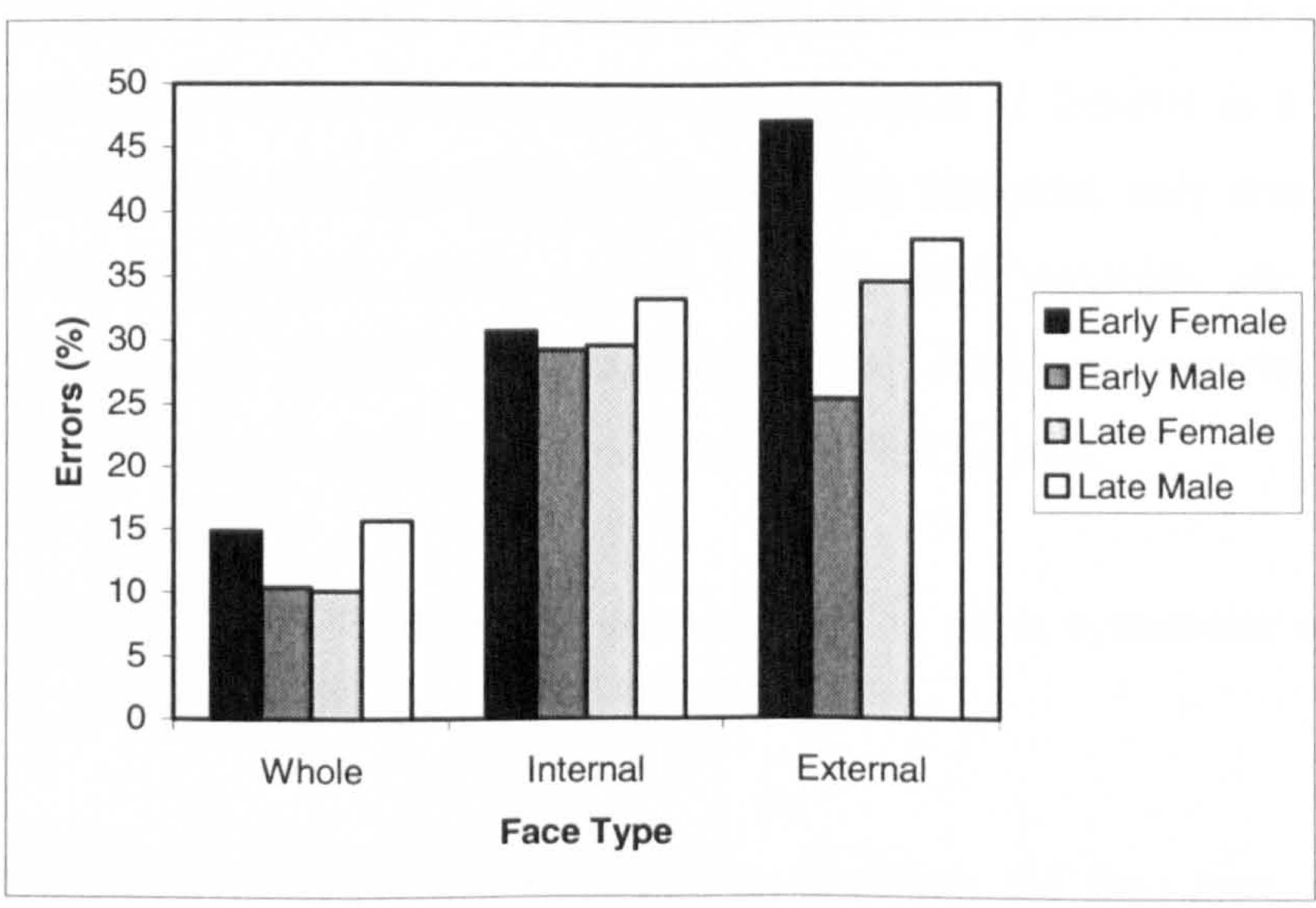
Figure 5.2. *Effect of AoA on Familiarity Decision RTs as a Function of Gender (Experiment 5).*



Errors. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subject factors, and by-items with face type, AoA and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F1(2,70) = 42.59$, MSe = 14.59, $p<.001$; $F2(2,192) = 29.08$, MSe = 45.47, $p<.001$, confirming the observed tendency for wholes (mean n = 2.2, 12.8%) to be recognized more accurately than internals (mean n = 5.2, 30.7%), and internals more accurately than externals (mean n = 6.2, 36.2%). Tukey tests (HSD = 2.2) revealed that whole faces were recognised with significantly fewer errors than internal and external features, but the difference between internal and external features was not significant. The main effect of AoA was not significant in either analysis, F1(1,35) = .23, MSe = 3.67, p = .64; F2(1,192) = .04, MSe = 45.47, p = .84, neither was the main effect of gender, F1(1,35) = 2.59, MSe = 7.73, p = .12; F2(1,192) = .93, MSe = 45.47, p = .34.

A two-way interaction between AoA and gender was significant in both analyses, $F_1(1,35) = 49.72$, MSe = 2.79, $p<.001$; $F_2(1,192) = 6.47$, MSe = 294.24, $p = .01$. In the by-subject analysis alone, the interaction between face type and gender also reached significance, $F_1(2,70) = 10.25$, MSe = 3.42, $p<.001$. These by-subjects interactions were subsumed in a significant three-way interaction between face type, AoA and gender, $F_1(2,70) = 13.42$, MSe = 2.09, $p<.001$, see Figure 5.3. As a means of refining the extent of the AoA effect, data from each face type were submitted independently to 2 x 2 within-subjects ANOVAs with AoA and gender as factors.

Figure 5.3. *Relationship Between AoA and Gender as a Function of Face Type, Indexed by Familiarity Decision Accuracy (Experiment 5).*



In the whole face condition, neither AoA, $F_1(1,35) = 0.06$, MSe = 0.99, $p = .80$, nor gender, $F_1(1,35) = 0.11$, MSe = 3.04, $p = .74$, approached significance. However, the interaction was significant, $F_1(1,35) = 15.10$, MSe = 1.71, $p<.001$. Tukey tests (HSD = 0.8) revealed that whilst male faces were affected by AoA, females were not affected. Further, late acquired females received fewer errors than late acquired males, but there was no effect of gender on early acquired faces. In the analysis of internals, AoA, $F_1(1,35) = 0.52$, MSe = 3.85, $p = .48$, gender, $F_1(1,35) = 0.32$, MSe = 3.65, $p = .57$, and the interaction, $F_1(1,35) = 2.54$, MSe = 2.63, $p = .12$, failed to reach significance. Whilst an analysis of the externals revealed no main effect of AoA, $F_1(1,35) = 0.01$, MSe = 2.91, $p = .96$, gender was significant, $F_1(1,35) = 11.24$, MSe = 7.89, $p<.01$, and entered into a significant interaction with AoA, $F_1(1,35) = 61.72$, MSe = 2.63, $p<.001$. Tukey tests (HSD = 1.0) revealed an effect of AoA on male and female faces, although this effect countered expectation for females. Further, gender affected early faces with fewer errors made to males compared with females but late acquired faces were not affected by gender. No further interactions emerged from the analyses.

5.2.3.1 Participant Gender Analysis

A post-hoc analysis of the effect of participant gender was carried out as an attempt to shed light on the discrepant results associated with face gender. Reaction time and error data were submitted to mixed-design by-subject ANOVAs with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors and participant gender (male or female) as a between-subjects factor. For this, and subsequent subsidiary analyses, only results which indicate some systematic effect on the experimental variables are reported. Further, only those results with implications for AoA are reported in full. Systematic effects on face type and gender are reported in brief.

Reaction Times. Participant gender failed to exert systematic effects on face type, AoA or face gender.

Errors. Participant gender failed to modulate the face type and AoA variables, but it interacted with face gender. In short, there was no effect of participant gender on male face recognition accuracy, however female participants recognised female faces with significantly fewer errors than male participants. Further, female participants were not affected by face gender, whereas male participants recognised male faces significantly more accurately than female faces.

5.2.3.2 Presentation Order Analysis

A collection of subsidiary ANOVAs explored the impact of presentation order on each face type separately, to check whether or not presentation order exerted any systematic effects. Reaction time and error data were submitted to by-subjects analyses with AoA (early or late) and gender (male or female) as within-subject factors and presentation order (first, second or third) as a between-subjects factor, and by-items analyses with AoA and gender as between-subjects factors and presentation order as a within-subjects factor.

Reaction Times. Presentation order failed to modulate effects of AoA or gender in any of the face type conditions.

Errors. Presentation order failed to modulate effects of AoA or gender in any of the face type conditions.

5.2.3.3 Analysis of Second Presentation Groups

. The variable 'presentation order' was probed further to investigate the impact of being preceded by one face type compared with another. Block 2 data were subject to scrutiny but data from faces seen for the first or third time were not analysed because there are unlikely to be uneven order effects imparted on these presentations. Reaction time and error data for each face type were independently submitted to mixed-design ANOVAs. In the whole face condition, for example, data were analysed by-subjects with AoA (early or late) and gender (male or female) as within-subject factors and group (IWE or EWI) as a between-subjects factor.

Reaction Times. Group failed to modulate effects of AoA or gender in the whole and external conditions. However, an interaction between gender and group emerged in the internals analysis. In short, internals preceded by whole faces were recognized faster than those preceded by externals, regardless of gender. Moreover, male internals were recognised faster than female internals when preceded by the externals, but when internal features had been preceded by whole faces, there was no effect of gender. AoA effects were not modulated by 'group' in the internal condition.

Errors. Group failed to modulate effects of AoA or gender.

5.2.3.4 Familiarity Analysis

Post-hoc analyses were carried out on the famous face and unfamiliar face data to assess the impact of familiarity on familiarity decisions. These analyses were carried out to provide a comparison with post-hoc familiarity analyses of gender classification performance (Experiment 6).

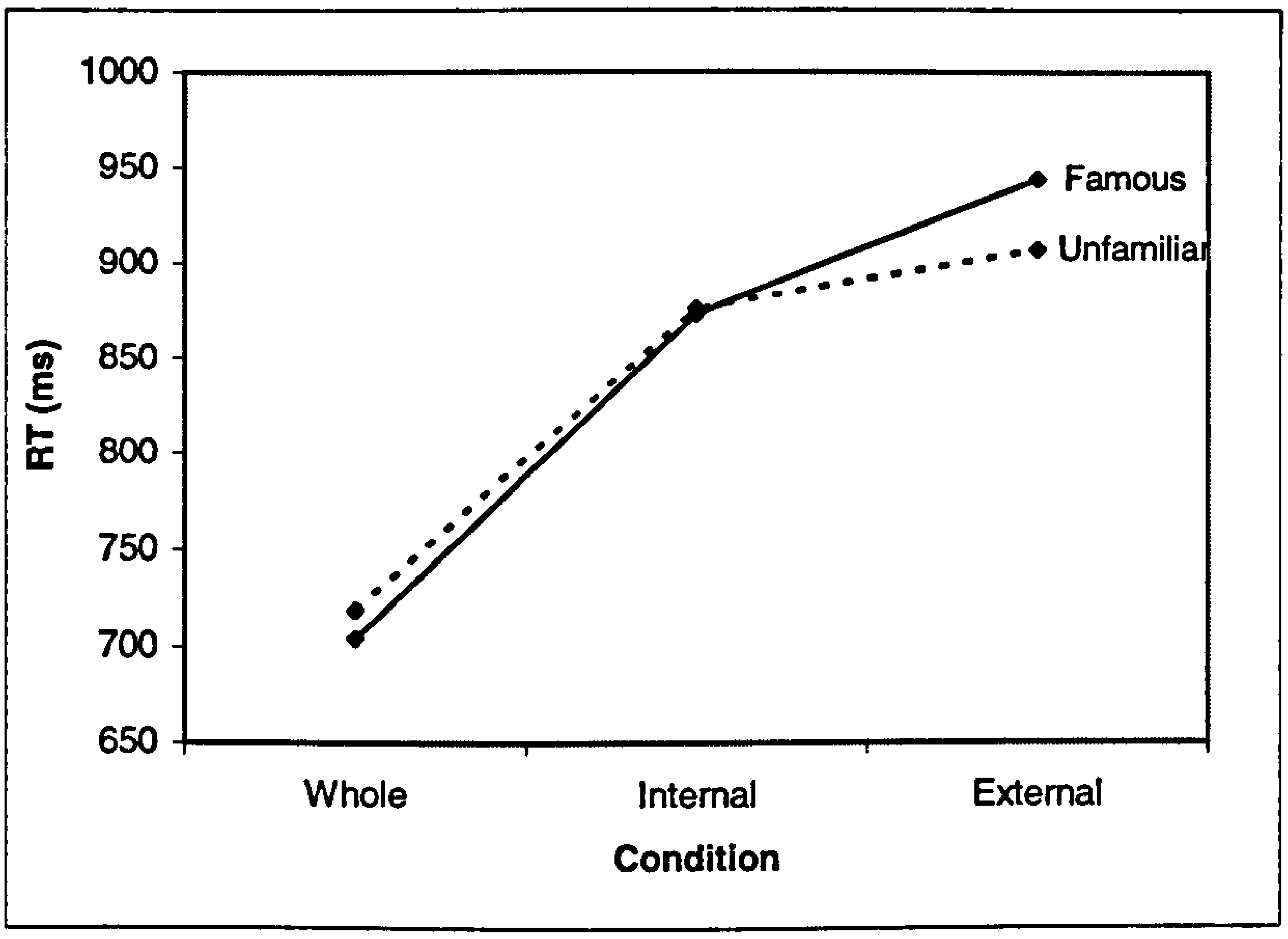
For the unfamiliar faces, the error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions made to faces correctly identified as 'unfamiliar' in the familiarity verification session; b) a failure to correctly identify a face as 'unfamiliar' in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect familiarity decisions, 45 (1.8%) whole face responses, 122 (5.0%) internal feature responses, and 213 (8.7%) external feature responses, were removed. In addition, 70 (2.9%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. Further, 43 (1.8%) responses from the whole condition were deemed outliers, as were 43 (1.8%) internal condition responses and 39 (1.6%) external.

Only correct familiarity decisions, and those made to faces correctly classified as 'unfamiliar' in the familiarity verification session, were submitted to the latency analysis. See Table 5.2 for mean famous and unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous or unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,70) = 45.73$, MSe = 39,271.64, $p<.001$; $F_2(2,396) = 272.02$, MSe = 6,239.36, $p<.001$, with increasing familiarity decision latencies to wholes (712 ms), internals (874 ms) and externals (926 ms). Tukey tests ($\alpha = .05$) revealed that whole faces were classified significantly faster than internal and external features, but the difference between the latter conditions was not significant (HSD = 112). Familiarity was not significant in either analysis, $F_1(1,35) = 0.43$, MSe = 9,752.84, $p = .52$; $F_2(1,396) = 2.65$, MSe = 6,239.36, $p = .11$. The main effect of gender was significant in both analyses, $F_1(1,35) = 41.69$, MSe = 3,384.11, $p<.001$; $F_2(1,396) = 12.76$, MSe = 6,239.36, $p<.001$, stemming from the faster classification of male faces (819 ms) compared with females (855 ms).

A two-way interaction between face type and familiarity approached significance in the by-subjects analysis, F1(2,70) = 2.86, MSe = 9,500.73, p = .064, attaining significance by-items, F2(2,396) = 6.85, MSe = 6,239.36, p = .001, see Figure 5.4. A simple main effects analysis revealed no effects of familiarity on whole faces, F1(1,35) = 2.62, MSe = 1,586.02, p = .11, internals, F1(1,35) = 0.03, MSe = 5,179.30, p = .86, or externals, F1(1,35) = 3.28, MSe = 7,597.71, p = .08. Further, effects of face type were shown to be significant for famous, F1(2,70) = 41.47, MSe = 13,192.08, p<.001, and unfamiliar faces, F1(2,70) = 32.52, MSe = 11,194.99, p<.001. For famous (HSD = 79) and unfamiliar faces (HSD = 73) Tukey tests attributed this significance to a difference between wholes and the other face types, with the internal and external conditions not differing significantly.

Figure 5.4. *Effect of Familiarity on Familiarity Decision RTs as a Function of Face Type (Experiment 5).*

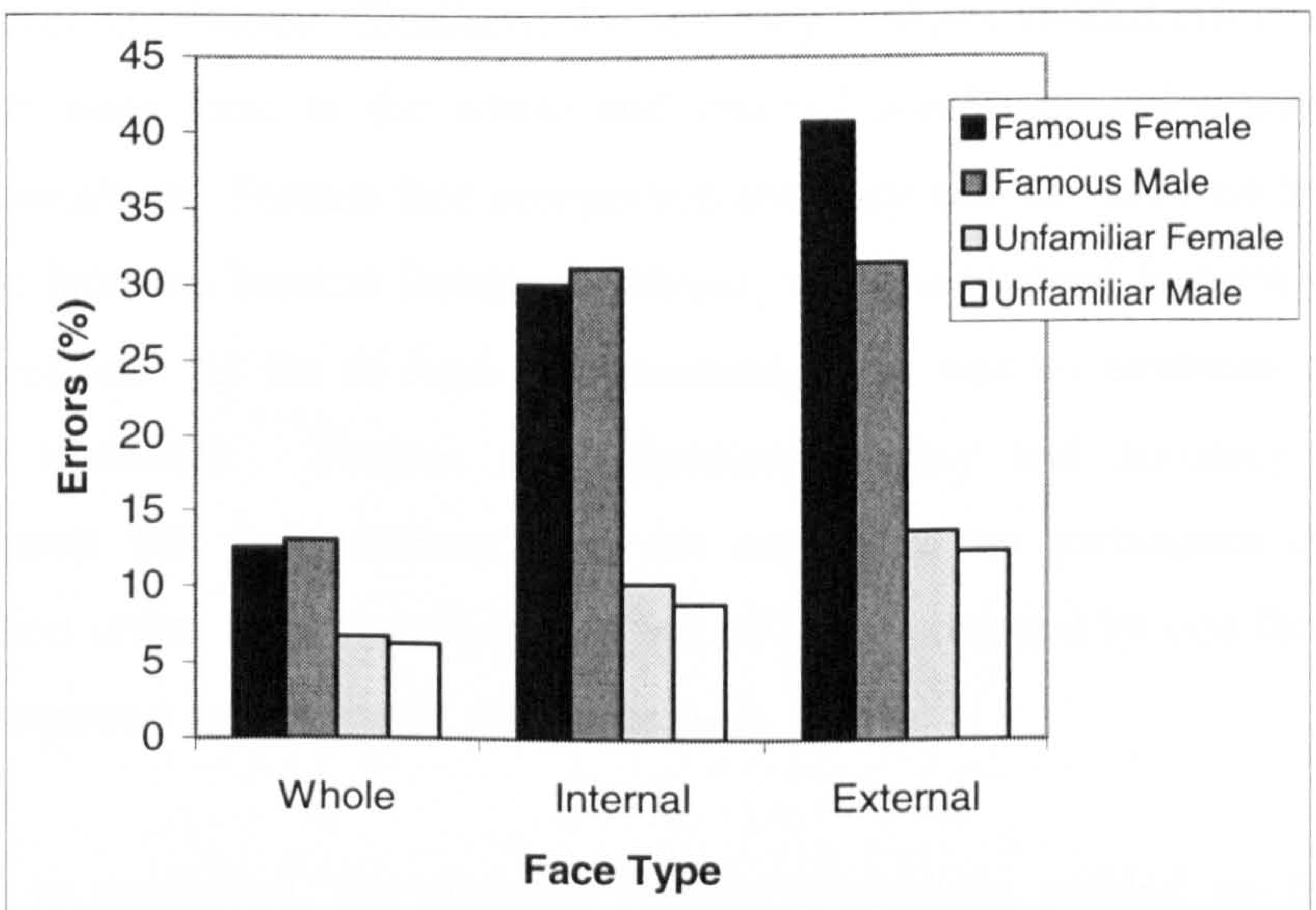


Errors. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous or unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,70) = 62.38$, MSe = 15.84, $p<.001$; $F_2(2,396) = 36.56$, MSe = 28.79, $p<.001$, with increasing errors to wholes (mean n = 3.3, 9.6%), internals (mean n = 6.9, 20.2%) and externals (mean n = 8.4, 24.7%). Tukey tests (HSD = 2.3) revealed that whole faces were classified with significantly fewer errors than internal and external features, but the difference between the latter conditions was not significant. The main effect of familiarity was significant in both analyses, $F_1(1,35) = 59.26$, MSe = 59.58, $p<.001$; $F_2(1,396) = 129.63$, MSe = 28.79, $p<.001$, with unfamiliar faces (mean n = 3.3, 9.7%) classified more accurately than famous faces (mean n = 9.0, 26.5%). The main effect of gender was significant by-subjects, $F_1(1,35) = 5.28$, MSe = 7.76, $p<.05$, with male faces (mean n = 5.9, 17.2%) classified more accurately than females (mean n = 6.5, 19.0%). This effect was not significant by items, $F_2(1,396) = 1.53$, MSe = 28.79, $p = .22$.

A significant two-way interaction between familiarity and face type emerged from both analyses, $F_1(2,70) = 16.53$, MSe = 20.99, $p<.001$; $F_2(2,396) = 12.82$, MSe =

28.79, $p<.001$. In the by-subject analysis, an interaction between face type and gender also reached significance, $F_1(2,70) = 8.38$, MSe = 4.57, $p = .001$. These by-subjects interactions were subsumed in a significant three-way interaction between face type, familiarity and gender, $F_1(2,70) = 7.47$, MSe = 4.35, $p = .001$, see Figure 5.5. As a means of refining the extent of the AoA effect, the data from each face type were submitted independently to 2 x 2 within-subjects ANOVAs with familiarity and gender as factors.

Figure 5.5. *Relationship Between Face Type and Gender as a Function of Familiarity, Indexed by Familiarity Decision Accuracy (Experiment 5).*



In the whole condition, the main effect of familiarity was significant, $F_1(1,35) = 21.16$, MSe = 7.88, $p<.001$, with more errors made classifying famous faces compared with unfamiliar faces. Gender was not significant, $F_1(1,35) = 0.01$, MSe = 3.48, $p = .97$, and neither was the interaction, $F_1(1,35) = 0.28$, MSe = 4.25, $p = .60$. For internal features, familiarity was significant, $F_1(1,35) = 47.01$, MSe = 39.33, $p<.001$, with famous internals attracting more errors than unfamiliar internals. Gender was not significant, $F_1(1,35) = 0.03$, MSe = 4.21, $p = .87$, and neither was the interaction, $F_1(1,35) = 0.79$, MSe = 7.89, $p = .38$. For external features, the main effects of familiarity, $F_1(1,35) = 40.64$, MSe = 54.36, $p<.001$, and gender, $F_1(1,35) = 12.75$, MSe = 9.20, $p = .001$, were significant, as was the interaction, $F_1(1,35) = 6.50$, MSe = 9.84, $p<.05$. Tukey tests (HSD = 1.99) demonstrated that familiarity affected male and female faces alike, with famous faces classified less accurately than unfamiliar faces. Whilst gender did not modulate the classification of unfamiliar faces, famous female faces attracted more errors than famous male faces.

5.2.4 Discussion

The latency analysis revealed customary AoA effects on wholes, internal features and external features, but these were restricted to male faces. Early acquired female faces were classified sufficiently slowly to equate latencies in the early and late acquired conditions. Similarly, the accuracy analysis yielded customary AoA effects on male faces in the whole and external conditions and a trend in the internal condition. Female face recognition accuracy was not affected by AoA in the whole face and internal feature conditions, whereas external features showed a reversed effect. As far as AoA is concerned, there was no evidence of speed-accuracy trade-offs. Further, the subsidiary latency and accuracy analyses demonstrated that AoA effects were not modulated by participant gender or presentation order. Additionally, the effect of being preceded by one face type or another imparted no systematic effects on AoA.

Contrary to prediction, the post-hoc familiarity analysis yielded no familiarity effects on response latencies, although there was a trend in the external condition for famous faces to be recognised more slowly than unfamiliar faces. This signalled possible interference from familiarity processes, for example, an additional level of processing to identify to whom the external features belonged to. The null effect in the latency analysis was not surprising given that it is suboptimal to compare responses to categories corresponding to separate binary decisions. Further, famous and unfamiliar sets were not matched on factors such as distinctiveness, which could have adversely affected the familiarity analysis. More errors were made to famous faces compared with unfamiliar faces, as anticipated. These results were consistent across face type.

It is surprising that AoA effects were restricted to male faces as there were no *a priori* reasons to expect discrepancies associated with gender. The result could be an artefact of the particular set of famous faces used. Nevertheless, this discrepancy is taken into account when interpreting findings from the forthcoming gender classification task: it is unlikely that AoA will affect female faces, given that it failed to influence familiarity decisions to these faces in the first instance.

5.3 Experiment 6

5.3.1 Introduction

Experiment 6 replicated Experiment 5 as a gender classification task with gender decisions made to early and late acquired whole faces, internal and external features. As an improvement on the earlier gender decision task (Experiment 4), a familiarity verification session administered after the experiment ensured that the latency analysis comprised only those responses to faces known by participants.

AoA was hypothesised to affect gender classifications made to internal features. In the whole and external feature conditions, it was hypothesised that performance would not be modulated by AoA because superficial cues to gender can facilitate these decisions. In light of the results of Experiment 5, it was hypothesised that any AoA effects would be confined to the classification of male faces. It is reasonable to suppose that null effects associated with female faces in Experiment 5 are translated into null effects in Experiment 6.

In addition, an effect of familiarity was predicted to emerge from the post-hoc analysis, with famous faces classified according to gender faster and more accurately than unfamiliar faces. More specifically, this effect was predicted to emerge with internal features only.

5.3.2 Method

Participants. Thirty six undergraduate students from the University of York took part in this experiment (8 male, 28 female), receiving course credit or a payment of £2 for their participation. Participants were 18 – 21 years old (mean = 19.67; S.D. = 1.01) and were required to have lived in the UK for at least 18 years.

Design. The experimental variables comprised face type (with three levels; whole face, internal features or external features), AoA (with two levels; early or late) and gender of face (with two levels; male or female) which were explored in a repeated measures design. The order of the face type conditions was counterbalanced across participants yielding six different predefined orders of presentation.

The dependent variables, response latency and accuracy, were the products of a gender classification task.

Materials. See Experiment 5 details, as the same stimuli were employed in this gender classification task, p. 141.

Procedure and Apparatus. See Experiment 5 details, p. 143. Note, however, that during the experiment proper participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is male or female".

5.3.3. Results

All participants met the criteria which required them to: a) produce 75% correct responses to famous faces in the whole face condition and familiarity verification session; b) produce 75% correct responses to famous faces and unfamiliar faces pooled together, in the whole face condition and familiarity verification session; c) produce 75% correct responses to unfamiliar faces in the internal and external conditions.

The error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications made to known faces (a face is 'known' if it is successfully recognized in the familiarity verification session); b) a failure to recognise a face in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect gender classifications made to known faces, 74 (3.0%) whole face responses, 94 (3.8%) internal feature responses, and 96 (3.9%) external feature responses were removed. In addition, 121 (4.9%) responses from each face type condition were removed as a consequence of faces not being recognized in the familiarity verification session. As a result of eliminating outliers slower than 3 standard deviations away from the mean and faster than 300ms, 67 (2.7%) whole face responses were removed, as were 41 (1.7%) from the internal condition and 50 (2.0%) from the external condition. Outliers were calculated for each face type separately.

Only correct gender classifications, and those made to faces successfully recognized in the familiarity verification session, were submitted to the latency analysis. Table 5.3 shows the mean correct response latencies in each condition of the experiment in addition to associated error rates. For inspection purposes, Table 5.3 additionally reports '% misclassifications' which denotes incorrect gender classifications made to faces classified correctly during the familiarity verification session.

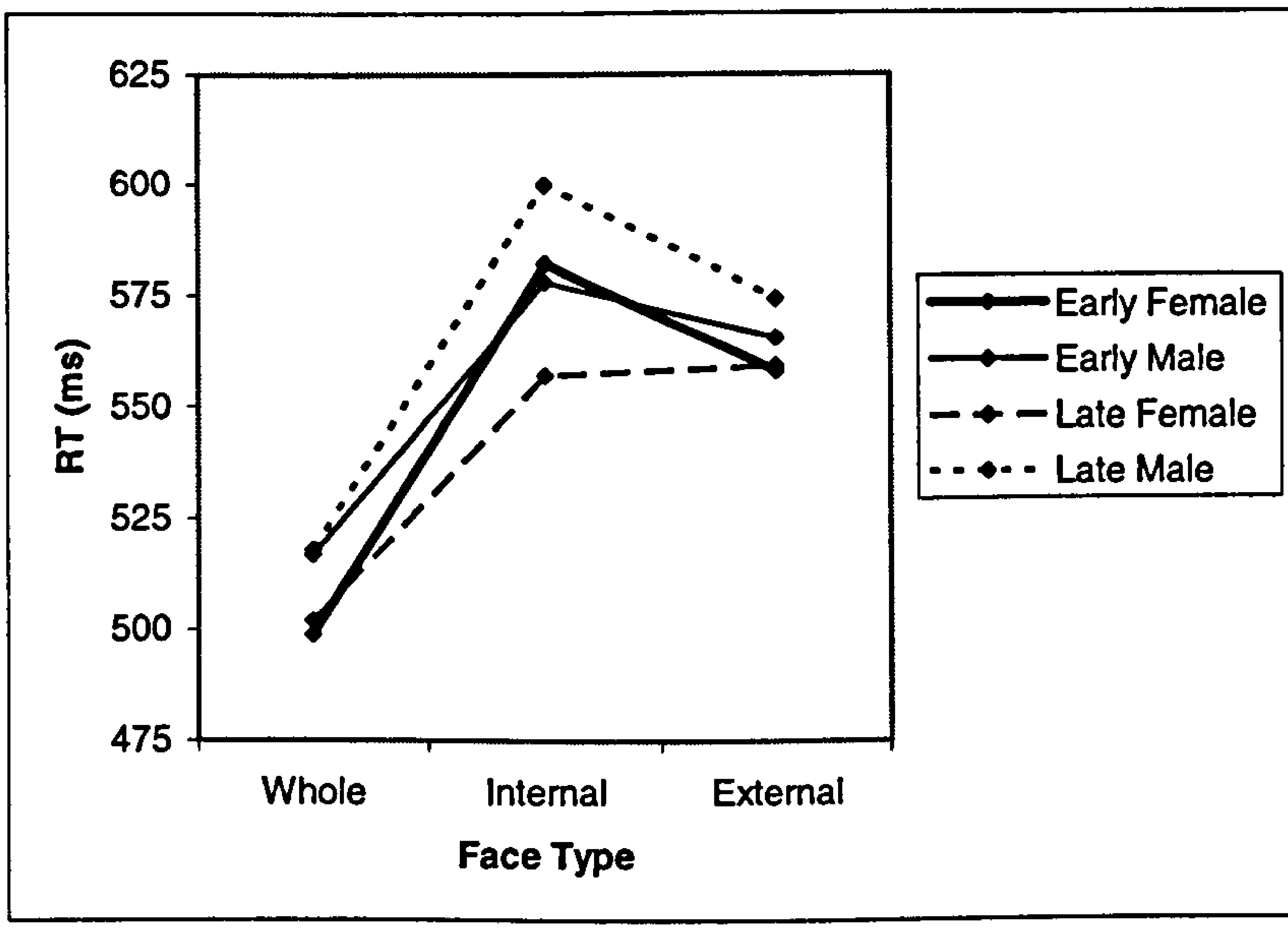
Table 5.3. *Mean RTs and Percentage Errors for Gender Decisions in Each Condition (Experiment 6).*

Stimulus		Early Acquired			Late acquired			Unfamiliar		
		Male	Female	Mean	Male	Female	Mean	Male	Female	Mean
Whole	*Mean RT*	517	499	508	518	502	510	502	518	510
	SD	78	76	74	70	74	69	67	66	65
	% Error	13.1	8.8	10.9	13.9	7.0	10.5	7.8	13.2	10.5
	% Misclass.	5.1	2.3	3.7	3.3	2.1	2.7	2.4	3.3	2.8
Internal	*Mean RT*	578	582	580	600	557	578	591	602	596
	SD	78	88	76	88	85	81	94	105	91
	% Error	11.4	9.5	10.5	14.7	6.2	10.5	7.1	16.8	11.9
	% Misclass.	4.7	3.4	4.1	5.6	2.8	4.2	2.7	6.9	4.8
External	*Mean RT*	565	558	560	574	559	566	575	606	589
	SD	64	87	64	80	72	68	78	93	74
	% Error	13.9	8.2	11.0	14.9	6.7	10.8	7.1	20.5	13.8
	% Misclass.	6.7	2.1	4.4	4.9	2.8	3.8	3.3	10.4	6.8

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors, and by-items with face type, AoA and face gender as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of face type was significant in both analyses, $F_1(2,70) = 30.61$, MSe = 6,422.60, $p<.001$; $F_2(2,192) = 39.52$, MSe = 2,314.29, $p<.001$, with wholes (509 ms) classified according to gender more quickly than externals (564 ms), and externals more quickly than internals (579 ms). With alpha set at .05, Tukey tests revealed that whole faces were classified significantly faster than internal and external features, but the classification latency of the latter two conditions did differ significantly (HSD = 45). The main effect of AoA was not significant in either analysis, $F_1(1,35) = 0.40$, MSe = 1,084.46, $p = .53$; $F_2(1,192) = 0.40$, MSe = 2,314.29, $p = .84$. However, the main effect of gender was significant in both analyses, $F_1(1, 35) = 11.23$, MSe = 2,515.16, $p<.01$; $F_2(1,192) = 9.97$, MSe = 2,314.29, $p<.01$, stemming from the faster classification of female faces (543 ms) compared with males (559 ms).

A significant two-way interaction between AoA and gender emerged from the by-subjects analysis, $F_1(1,35) = 6.41$, MSe = 1,251.20, $p<.05$. This interaction was incorporated into a three-way interaction between face type, AoA and gender, $F_1(2,70) = 4.89$, MSe = 1,228.03, $p = .01$, see Figure 5.6. Data from each face type condition were independently scrutinized in 2 x 2 within-subject ANOVAs to probe the effect of AoA on gender.

Figure 5.6. *Relationship Between Gender and AoA as a Function of Face Type, Indexed by Gender Decision RTs (Experiment 6).*

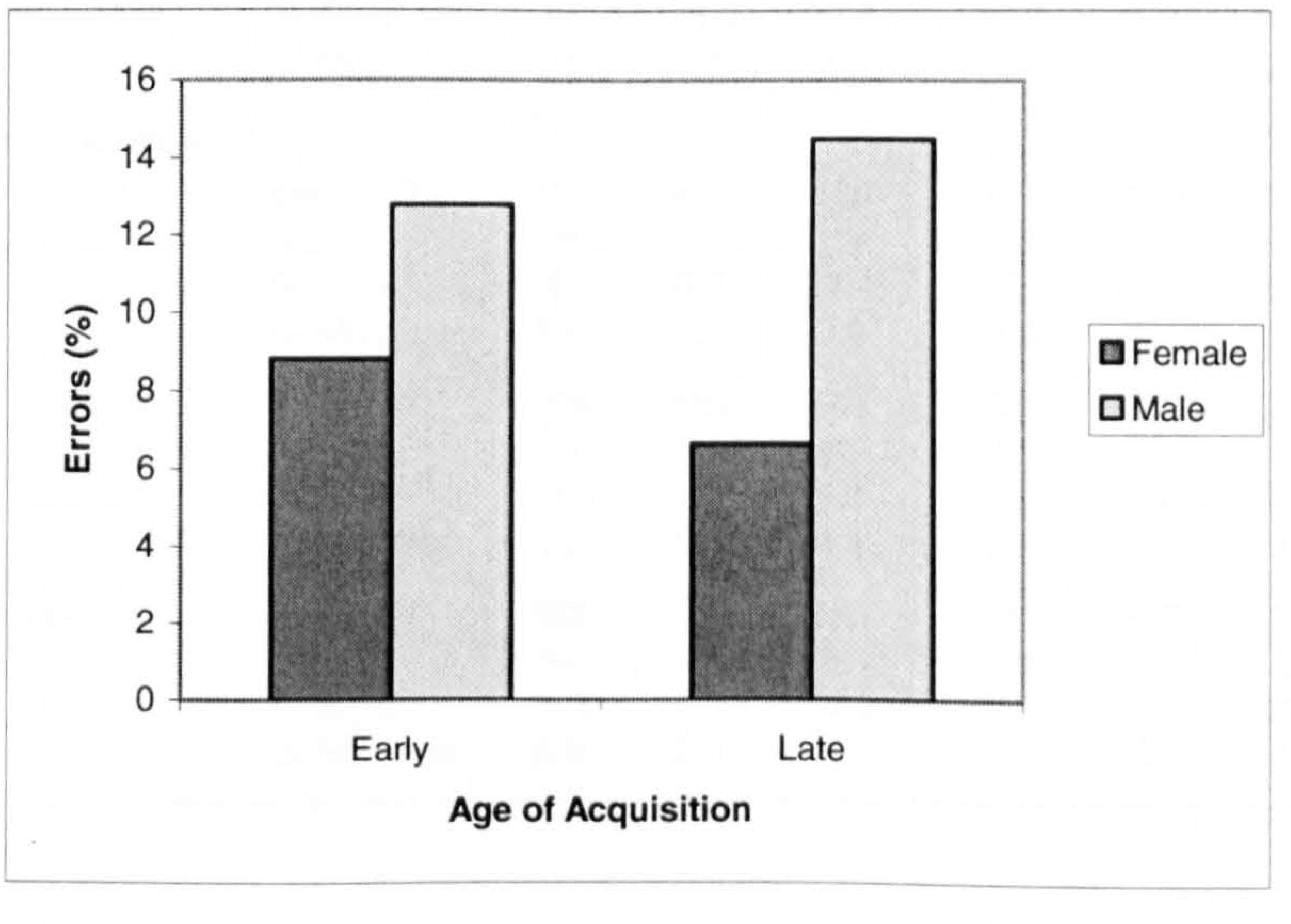


In the whole face condition, whilst the main effect of gender was significant, $F_1(1,35) = 11.20$, MSe = 992.55, $p<.01$, AoA was not significant, $F_1(1,35) = 0.24$, MSe = 706.23, $p = .63$, and neither was the interaction, $F_1(1,35) = 0.05$, MSe = 890.00, $p = .82$. In the analysis of internal features, the main effect of gender was significant, $F_1(1,35) = 5.42$, MSe = 2,600.78, $p<.05$, but AoA failed to reach significance, $F_1(1,35) = 0.40$, MSe = 833.34, $p = .83$. The interaction was significant, $F_1(1,35) = 18.89$, MSe = 1,031.29, $p<.001$. Tukey tests revealed that the customary effect of AoA on male face classification and the reversed AoA effect found for female faces, were both significant (HSD = 20). Further, late acquired face classification latency was modulated by gender with female faces classified according to gender faster than male faces; early acquired faces were not affected by gender. In the external features analysis, gender, $F_1(1,35) = 1.14$, MSe = 3,937.08, $p = .29$, and AoA, $F_1(1,35) = 0.12$, MSe = 1,209.98, $p = .40$, failed to reach significance as did the interaction, $F_1(1,35) = 0.28$, MSe = 1,785.98, $p = .60$. No further interactions emerged from the analyses.

Errors. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subject factors, and by-items with face type, AoA and face gender as between-subjects factors. The main effect of face type was not significant in either analysis, $F_1(2,70) = 0.18$, MSe = 1.20, $p = .84$; $F_2(2,192) = 0.04$, MSe = 11.95, $p = .96$. Similarly, the main effect of AoA failed to attain significance by-subjects or by-items, $F_1(1,35) = 0.08$, MSe = 2.45, $p = .78$; $F_2(1,192) = 0.03$, MSe = 11.95, $p = .86$. The main effect of gender was significant by-subjects, $F_1(1,35) = 17.34$, MSe = 6.29, $p<.001$, with female faces (mean n = 1.3, 3.9 %) attracting fewer errors than male faces (mean n = 2.3, 6.8 %). By-items, gender was not significant, $F_2(1,192) = 19.32$, MSe = 11.95, $p<.001$.

In the by-subjects analysis, a two-way interaction between AoA and gender reached significance, $F_1(1,35) = 4.05$, MSe = 2.88, $p = .05$, see Figure 5.7. Tukey tests (HSD = 1.1) attributed the source of this significance to the difference between late acquired male and female face classification accuracy, with females classified according to gender more accurately than males.

Figure 5.7. *Effect of AoA on Gender Decision Accuracy, as a Function of Gender (Experiment 6).*



5.3.3.1 Participant Gender Analysis

A post-hoc analysis of the effect of participant gender was carried out as an attempt to shed light on the discrepant AoA results associated with face gender. Reaction time and error data were submitted to mixed-design by-subjects ANOVAs with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors and participant gender (male or female) as a between-subjects factor. Table 5.4 shows the mean correct response latencies, associated error rates and '% misclassification' rates for male and female participants, in each condition of the experiment.

Table 5.4. *Mean RTs and Percentage Errors for Gender Decisions in Each Condition, Split According to Participant Gender (Experiment 6).*

		Early Acquired			Late acquired		
Stimulus		Male	Female	Mean	Male	Female	Mean
Male Participants							
Whole	*Mean RT*	553	533	544	538	532	535
	SD	77	68	71	57	90	73
	% Error	7.4	11.0	9.2	9.6	8.8	9.2
	% Misclass.	2.3	2.3	2.4	0.8	2.3	1.6
Internal	*Mean RT*	609	661	632	646	619	632
	SD	89	122	99	103	105	95
	% Error	9.6	16.2	12.9	11.8	10.3	11.0
	% Misclass.	5.3	2.3	3.9	2.5	3.2	2.8
External	*Mean RT*	563	618	587	600	598	598
	SD	67	101	69	93	80	70
	% Error	9.6	15.4	12.5	10.3	8.8	9.6
	% Misclass.	5.5	1.6	3.5	3.2	1.5	2.3
Female Participants							
Whole	*Mean RT*	507	489	498	513	493	503
	SD	76	76	73	73	68	67
	% Error	14.7	8.2	11.4	15.1	6.5	10.8
	% Misclass.	5.6	2.4	3.9	4.5	2.1	3.2
Internal	*Mean RT*	569	559	564	587	540	562
	SD	73	61	63	81	71	71
	% Error	12.0	7.6	9.8	15.5	5.0	10.3
	% Misclass.	4.8	3.4	4.1	6.4	2.8	4.5
External	*Mean RT*	566	541	552	567	548	557
	SD	64	76	61	76	67	66
	% Error	15.1	6.1	10.6	16.2	6.1	11.1
	% Misclass.	6.9	2.4	4.6	5.5	3.2	4.3

Reaction Times. Participant gender failed to modulate face type. However, this factor entered into a significant two-way interaction with face gender, which was subsumed in a three-way interaction between participant gender, face gender and AoA which reached borderline significance, $F_1(1,34) = 3.77$, MSe = 1,159.37, $p = .06$. Male and female face data were independently scrutinized in 2 x 2 within-subject ANOVAs to probe the effect of participant gender on AoA. In the female face analysis, participant gender reached significance, $F_1(1,34) = 7.20$, MSe = 7,375.12, $p = .01$, with female participants classifying female faces faster than male participants. The main effect of AoA reached significance, $F_1(1,34) = 3.97$, MSe = 440.15, $p = .05$, but the direction of the trend countered expectation. The interaction was not significant, $F_1(1,34) = 2.57$, MSe = 440.15, $p = .12$. In male face analysis, participant gender was not significant, $F_1(1,34) = 1.79$, MSe = 7,717.07, $p = .19$, however AoA was significant, $F_1(1,34) = 7.45$, MSe = 317.81, $p = .01$, with early acquired faces classified faster than late acquired faces. The interaction was not significant, $F_1(1,34) = 1.33$, MSe = 317.81, $p = .26$. Importantly, AoA effects do not vary as a function of participant gender.

Errors. Participant gender failed to modulate face type and AoA, but interacted with face gender. In short, there was no effect of participant gender on male face classification accuracy, however female faces were classified with significantly fewer errors by female participants compared with male participants. Further, female participants classified female faces significantly more accurately than male faces, whereas male participants were not affected by face gender.

5.3.3.2 Presentation Order Analysis

A collection of subsidiary ANOVAs explored the effect of presentation order on each face type separately. Reaction time and error data were submitted to by-subjects analyses with AoA (early or late) and gender (male or female) as within-subject factors and presentation order (first, second or third) as a between-subjects factor, and by-items analyses with AoA and gender as between-subjects factors and presentation order as a within-subjects factor.

Reaction Times. Presentation order failed to modulate effects of gender in the whole and internal feature conditions. However, the by-subjects and by-items external feature analyses recovered an interaction between gender and order: female faces were classified faster than male faces during the first block, but no gender effects emerged during second and third presentations. Further, there was no effect of presentation order on female faces, whereas the effect of presentation order reached borderline significance for male faces, but when probed, no differences emerged over successive presentations. Presentation order failed to exert systematic effects on AoA regardless of face type.

Errors. Presentation order failed to modulate effects of gender in the whole and internal feature conditions. However, the by-items analyses of externals recovered an interaction between gender and order: female faces were classified more accurately than male faces during the first block, but no gender effects emerged during second and third presentations. Further, there was no effect of presentation order on female faces, whereas male faces were affected by presentation order with gender classification accuracy increasing over successive presentations. Presentation order failed to modulate effects on AoA regardless of face type.

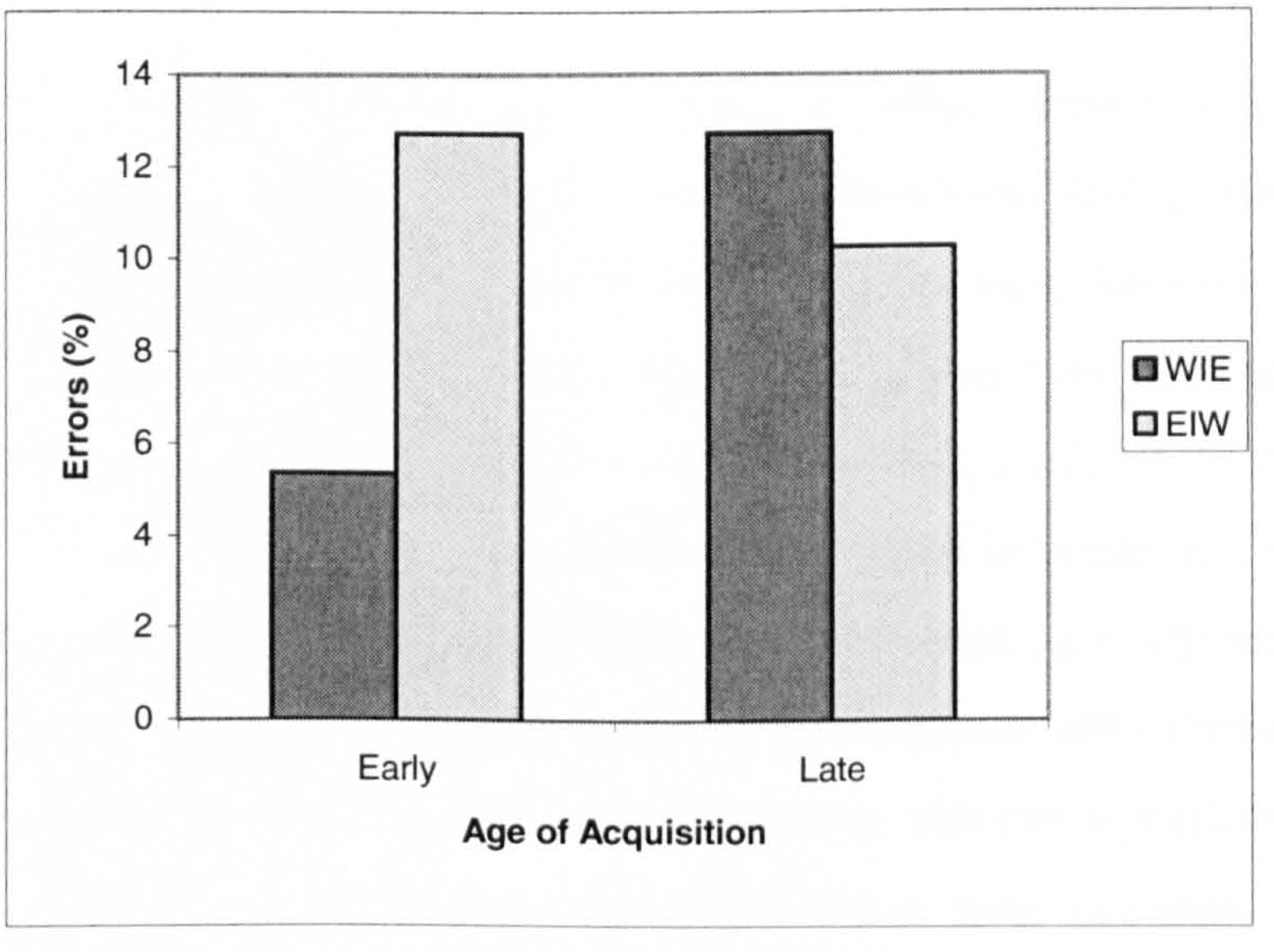
5.3.3.3 Analysis of Second Presentation Groups

The variable 'presentation order' was probed further to investigate the impact of being preceded by one face type compared with another. Block 2 data were subject to scrutiny. Reaction time and error data for each face type were independently submitted to mixed-design ANOVAs. In the whole face condition, for example, data were analysed by-subjects with AoA (early or late) and gender (male or female) as within-subject factors and group (IWE or EWI) as a between-subjects factor.

Reaction Times. Group failed to modulate effects of AoA or gender regardless of face type.

Errors. Group failed to modulate effects of AoA in the whole and external conditions, but the analysis of internals yielded an interaction between AoA and group which reached borderline significance, $F_1(1,10) = 4.15$, MSe = 2.01, $p = .069$, see Figure 5.8. Tukey tests (HSD = 2.5) indicated no significant differences between the cells of the design. Although group consistently failed to modulate effects of AoA and gender, the pattern in the internals analysis indicates that AoA is systematically affected as a consequence of being preceded by one face type or another.

Figure 5.8. *Effect of AoA on Gender Decision Accuracy, as a Function of Group (Experiment 6).*



5.3.3.4 Familiarity Analysis

Famous face and unfamiliar face data were analysed together to provide an additional tool with which to assess the notion that gender classifications are carried out independently of identity information.

For the unfamiliar faces, the error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications made to faces correctly identified as 'unfamiliar' in the familiarity verification session; b) a failure to correctly identify a face as 'unfamiliar' in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect gender classifications, 64 (2.6%) whole face responses, 114 (4.7%) internal feature responses, and 164 (6.7%) external feature responses, were removed. In addition, 131 (5.4%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. Further, 61 (2.5%) responses from the whole condition were deemed outliers, as were 47 (1.9%) internal condition responses and 43 (1.8%) external.

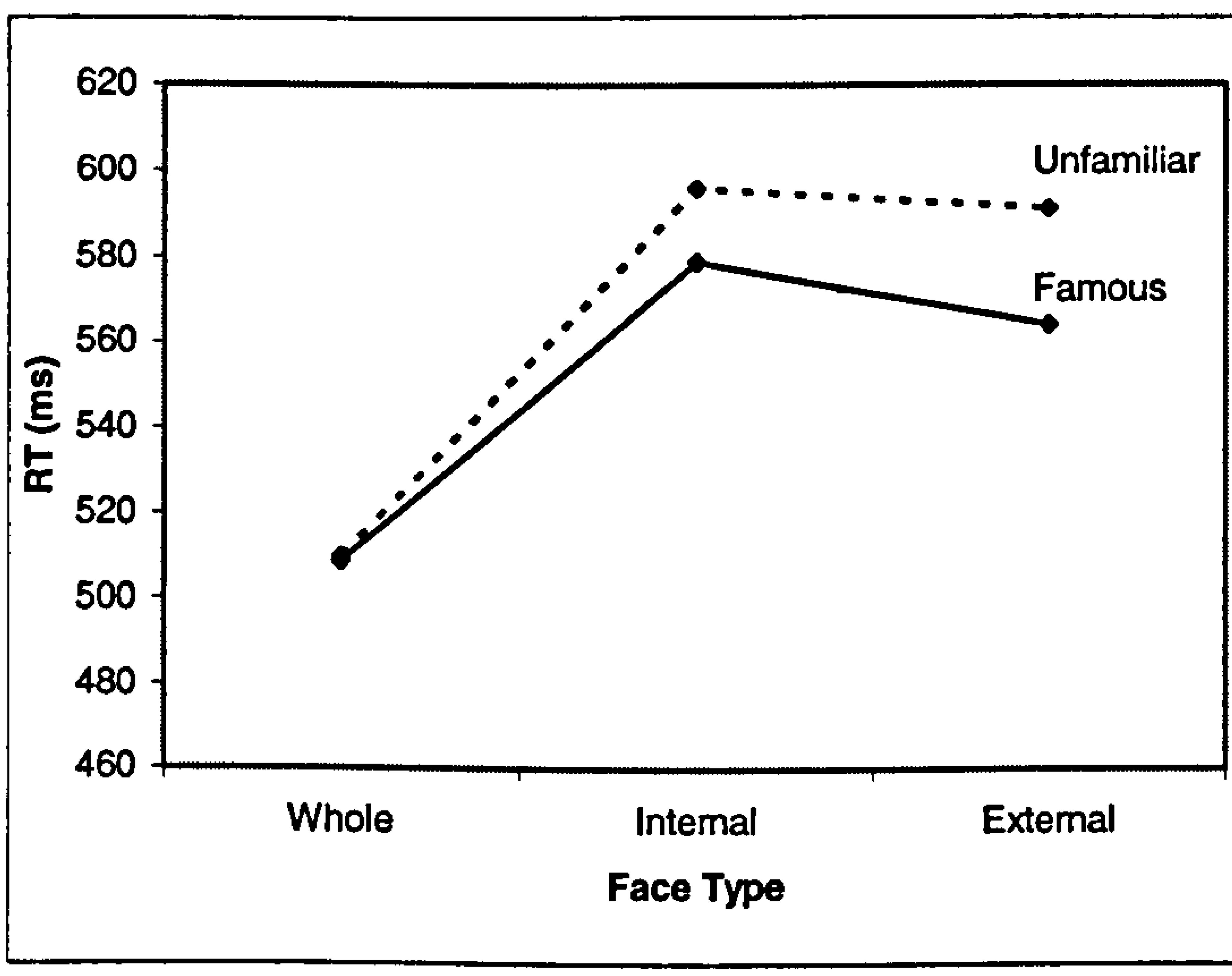
Only correct gender classifications and those made to faces correctly classified as 'unfamiliar' in the familiarity verification session, were submitted to the latency analysis. See Table 5.3 for mean unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous and unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,70) = 36.20$, MSe = 7,178.12, $p<.001$; $F_2(2,396) = 85.35$, MSe = 3,086.30, $p<.001$, with increasing gender classification latencies to wholes (509 ms), externals (577 ms) and internals (588 ms). Tukey tests ($\alpha = .05$) revealed that whole faces were classified significantly faster than internal and external features, but the difference between the latter two conditions was not significant (HSD = 48). The main effect of familiarity was significant in both analyses, $F_1(1,35) = 24.87$, MSe = 956.01, $p<.001$; $F_2(1,396) = 7.09$, MSe = 3,086.30, $p<.01$, stemming from the faster classification of famous faces (551 ms) compared with unfamiliar faces (566 ms). The main effect of gender was not significant in either analysis, $F_1(1, 35) = 0.07$, MSe = 3,090.63, $p = .79$; $F_2(1,396) = 0.02$, MSe = 3,086.30, $p = .88$.

A two-way interaction between face type and familiarity was significant in the by-subjects analysis, $F_1(2,70) = 8.07$, MSe = 759.75, $p = .001$, and reached borderline significance by-items, $F_2(2,396) = 2.92$, MSe = 3,086.30, $p = .06$, see Figure 5.9. A simple main effects analysis demonstrated that the classification of whole faces was not affected by familiarity, $F_1(1,35) = 0.04$, MSe = 235.27, $p = .84$, whereas internal, $F_1(1,35) = 11.05$, MSe = 486.07, $p<.01$, and external features, $F_1(1,35) =$

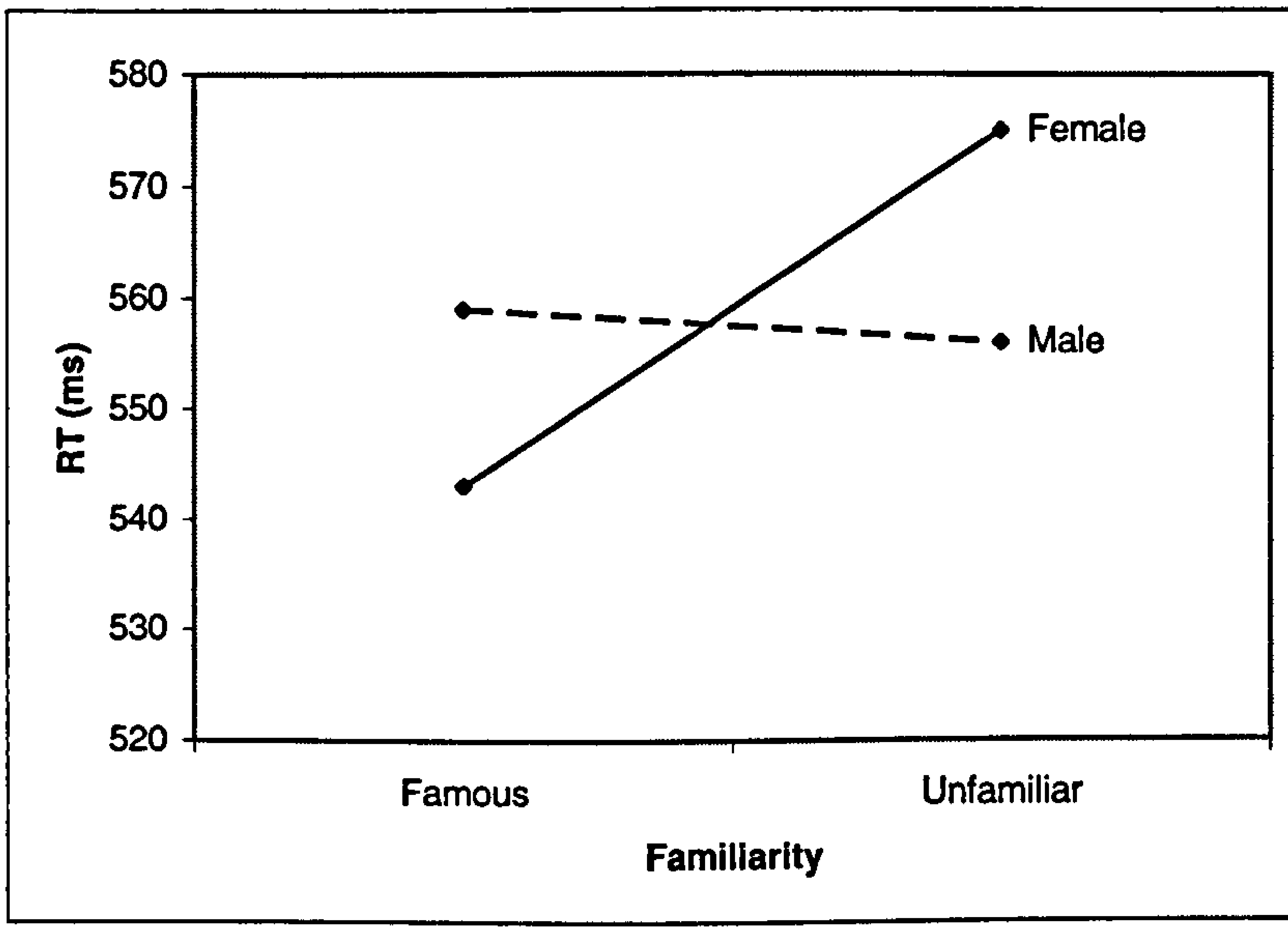
24.42, MSe = 513.24, p<.001, were affected by familiarity, with famous faces classified faster than unfamiliar faces. Moreover, famous face, $F_1(2,70) = 30.24$, MSe = 1,614.49, p<.001, and unfamiliar face classification performance, $F_1(2,70) = 35.64$, MSe = 2,359.28, p<.001, was modulated by face type. Tukey tests demonstrated that whole faces were classified faster than internal and external features with the latter two conditions not differing significantly. This was true for famous faces (HSD = 23) and unfamiliar faces (HSD = 27).

Figure 5.9. *Effect of Face Type on Gender Decision RTs as a Function of Familiarity (Experiment 6).*



A significant two-way interaction between familiarity and gender emerged from both analyses, $F_1(1,35) = 38.89$, MSe = 873.62, p<.001; $F_2(1,396) = 13.82$, MSe = 3,086.30, p<.001, see Figure 5.10. Tukey tests (HSD = 19) revealed that female faces were affected by familiarity, but male faces were not. Moreover, gender imparted no effect on famous faces, whereas male unfamiliar faces were classified faster than female unfamiliar faces.

Figure 5.10. *Effect of Familiarity on Gender Decision RTs as a Function of Gender (Experiment 6).*

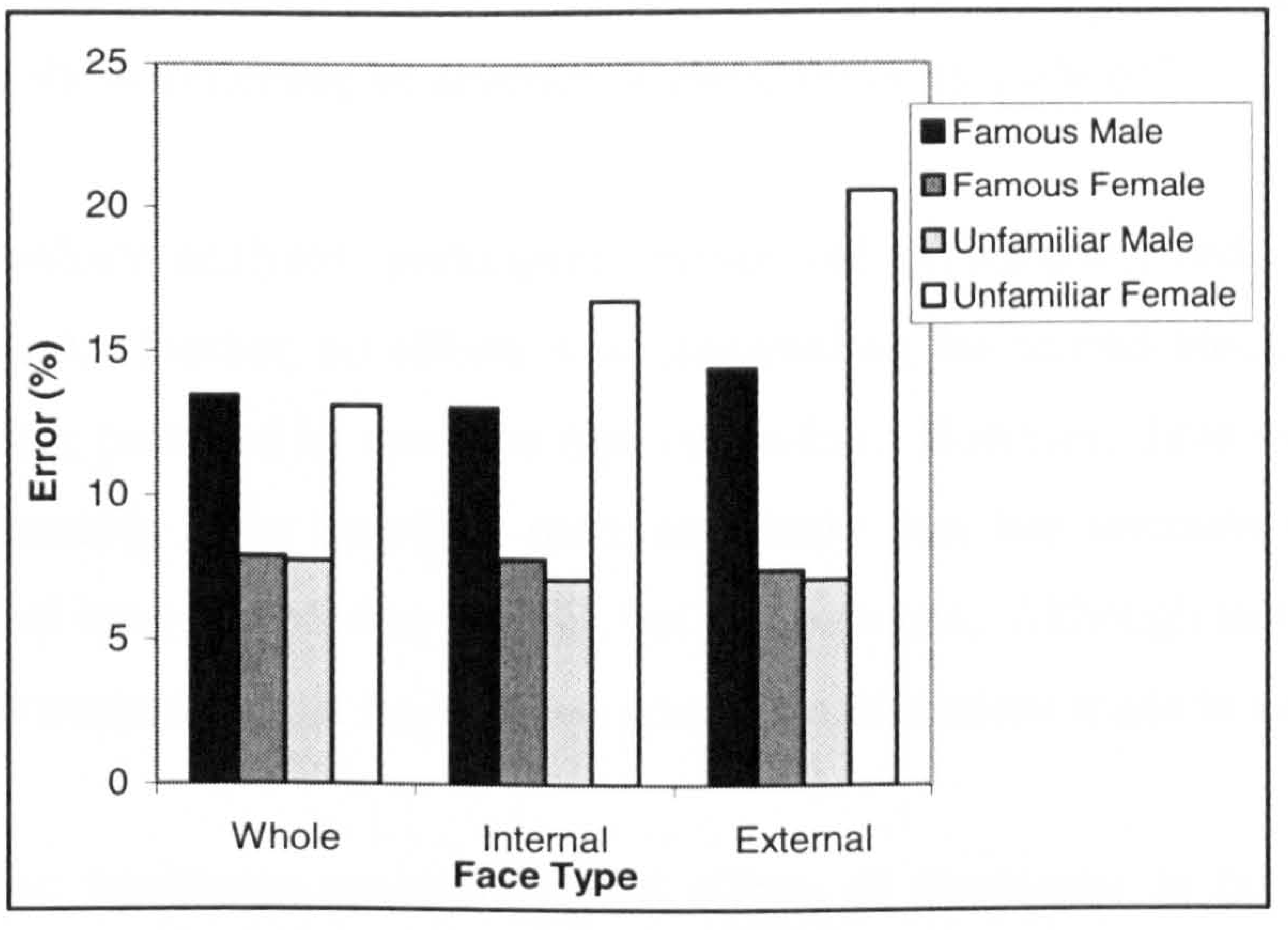


Errors. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous and unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant by-subjects, $F_1(2,70) = 4.89$, MSe = 2.77, $p = .01$, with increasing errors to wholes (mean n = 3.6, 10.6%), internals (mean n = 3.8, 11.2%) and externals (mean n = 4.2, 12.4%). Tukey tests (HSD = 0.9) indicated that there were no significant differences between the face type conditions. By-items, face type was not significant, $F_2(2,396) = 1.12$, MSe = 13.08, $p = .33$. The main effect of familiarity was not significant in either analysis, $F_1(1,35) = 2.33$, MSe = 10.13, $p = .14$; $F_2(1,396) = 1.80$, MSe = 13.08, $p = .18$. The main effect of gender was significant by-subjects, $F_1(1,35) = 4.77$, MSe = 8.33, $p<.05$, with male faces (mean n = 3.6, 10.6%) attracting fewer errors than female faces (mean n = 4.2, 12.4%). However, this was not significant by-items, $F_2(1,396) = 3.07$, MSe = 13.08, $p = .08$.

Significant two-way interactions between face type and familiarity, $F_1(2,70) = 5.64$, MSe = 1.83, $p<.01$, and face type and gender, $F_1(2,70) = 4.09$, MSe = 2.92, $p<.05$, emerged from the by-subjects analysis. Further, both analyses yielded significant interactions between familiarity and gender, $F_1(1,35) = 53.09$, MSe =

13.92, $p<.001$; $F_2(1,396) = 59.21$, MSe = 13.08, $p<.001$. The by-subjects interactions were subsumed in a significant interaction with face type, familiarity and gender, $F_1(2,70) = 18.57$, MSe = 1.25, $p<.05$, see Figure 5.11. Data from each face type were independently scrutinized in 2 x 2 within-subject ANOVAs to probe the effect of familiarity on gender.

Figure 5.11. *Relationship Between Gender and Familiarity as a Function of Face Type, Indexed by Gender Decision Accuracy (Experiment 6).*



In the whole face condition, neither familiarity, $F_1(1,35) = 0.07$, MSe = 3.75, $p = .80$, nor gender, $F_1(1,35) = 0.01$, MSe = 3.30, $p = .93$, were significant. However the interaction was significant, $F_1(1,35) = 20.04$, MSe = 6.22, $p<.001$. Tukey tests (HSD = 1.6) revealed that all conditions were significantly different from one another. In the internal condition, familiarity was not significant, $F_1(1,35) = 1.54$, MSe = 5.86, $p = .22$. However, gender was significant, $F_1(1,35) = 6.12$, MSe = 3.31, $p<.05$, and interacted significantly with familiarity, $F_1(1,35) = 41.51$, MSe = 5.54, $p<.001$. Tukey tests (HSD = 1.5) demonstrated all conditions to be significantly different from one another. Familiarity, $F_1(1,35) = 8.38$, MSe = 4.18, $p<.01$, and gender, $F_1(1, 35) = 5.73$, MSe = 7.57, $p<.05$, were significant in the external condition and interacted significantly, $F_1(1, 35) = 92.63$, MSe = 4.65, $p<.001$. Tukey tests (HSD = 1.37) indicated a complement of significant differences once again.

5.3.4 Discussion

As predicted, gender classification speed was affected by AoA in the internal condition. This effect was confined to male faces. Female internals displayed reverse AoA effects, having demonstrated no AoA effects in the previous familiarity decision task. It appears that where superficial cues to gender were present, no AoA effects emerged. Further, no effects of AoA emerged in the accuracy analysis, reflecting an absence of speed-accuracy trade-offs.

In the subsidiary analyses, participant gender and presentation order failed to modulate AoA. Further, no effects were imparted on the second block data as a result of being preceded by one face type or another. However, there was a trend for early internals to be classified more accurately than late internals when this condition had been preceded by wholes, but not externals. Although tentative, this supports the contention that AoA affects gender classifications made to internals.

The post-hoc familiarity analysis yielded effects of familiarity in the predicted direction. However, in addition to affecting internal features as predicted, famous externals were classified faster than unfamiliar externals. Although the effect of familiarity on internals supports Goshen-Gottstein and Ganel's (2000) post-hoc familiarity analysis, the unexpected effect of familiarity on externals somewhat undermines the result: familiarity effects should not arise when superficial cues facilitate gender decisions. Furthermore, familiarity effects were restricted to female faces. This is somewhat counter-intuitive because both AoA and familiarity effects should arise for the same reason and thus co-occur. The accuracy analysis yielded predicted familiarity effects, except these emerged across all face types. Once again, only female faces were affected by familiarity in the predicted direction; the effect reversed for male faces.

Discrepancies associated with face gender pervaded Experiment 5 and 6 with customary AoA effects restricted to male faces in both cases. Therefore, these experiments were replicated with different images to establish whether or not AoA effects are an exclusive property of male faces. Although unlikely, the literature does not document instances where male and female faces have been investigated separately. If AoA does affect familiarity decisions to male and female faces, this will provide a more robust baseline from which to tease out effects in the gender classification task.

5.3.4.1 Parallel-Route Versus Single-Route Explanations

It is theoretically possible to establish whether parallel-route or single-route accounts are most apt to explain the effect of AoA on gender decisions seen in the male face dataset. Parallel-route hypotheses predict such effects whenever processing via the visually-derived semantic route is sufficiently laborious. This could arise through androgyny or as a consequence of removing external cues to gender. According to this hypothesis, AoA could affect whole faces or internal features so long as these stimuli have relatively slow gender decision latencies and fast familiarity decisions. Stimuli with the converse response pattern are less likely to yield interactivity. Conversely, the single-route hypothesis predicts effects of identity on gender decisions when the perceptual whole is truncated; when external features are removed. Therefore, internals should be affected by AoA, regardless of the relative differences in the time-course of the operations and whole faces should not yield AoA effects under any circumstances.

In a post-hoc analysis, the difference between familiarity decision and gender decision latencies was calculated for wholes and internals. A median split of the difference scores produced a set of items with small difference scores (slow gender decisions and fast familiarity decisions) and a set of items with large difference scores (fast gender decisions and slow familiarity decisions). 'Difference score' was inserted as a factor into a repeated measures ANOVA with AoA and gender as factors. This analysis was carried out on the whole face and internal feature gender decision latencies from Experiment 6.

In the internal condition, the effect of AoA was not qualified by the relative time-courses of the operations. However, in the whole face analysis, 'difference score' entered into an interaction with AoA which reached borderline significance,

$F_1(1,35) = 3.89$, MSe = 1,857.80, $p = .057$. Whilst Tukey tests recovered no significant effects (HSD = 27), there was a trend approximating an AoA effect for items with relatively slow gender decisions and fast familiarity decisions. For these items, the difference between early and late acquired faces was 14 ms whereas faces with relatively fast gender decisions and slow familiarity decisions exhibited a smaller reversed AoA effect (6 ms).

The effect of 'difference score' on whole faces suggests that the parallel-route hypothesis may provide the most likely explanation for the relationship between identity and gender processing in Experiment 6. However, non-significant effects are not particularly persuasive. Moreover, this is not consistent with the null effect exhibited by internals, however it is possible that the effect of AoA on male internal features was sufficiently pronounced so as to preclude further qualification with another factor. Chapter 7 attempts to distinguish parallel-route and single-route explanations.

CHAPTER 6 – THE RELATIONSHIP BETWEEN IDENTITY AND GENDER PROCESSING: RECRUITING AOA AS AN INVESTIGATIVE TOOL - CONTINUED

6.1 Experiment 7

6.1.1 Introduction

Experiment 7 was carried out as a replication of Experiment 5 and a prerequisite to the forthcoming gender classification task (Experiment 8). Familiarity decisions were made to early acquired and late acquired whole faces, internal features and external features taken from the dated database. This database was used for pragmatic reasons: to replicate Experiments 5 and 6 with different stimuli to establish whether or not the various discrepancies associated with face gender were genuine phenomena or artefacts associated with the stimulus set. Errors in the internal and external conditions were high in Experiment 5, therefore the familiarity verification session was administered before the experiment proper, so that it could also serve as a practice session to improve recognition accuracy.

As before, AoA was hypothesised to affect familiarity decisions made to faces in all three face type conditions. Moreover, these effects were hypothesised to be consistent for male and female faces. In addition, the post-hoc familiarity analysis was hypothesised to yield familiarity effects with famous faces recognised faster than unfamiliar faces, and the converse trend in the accuracy analysis.

6.1.2 Method

Participants. Thirty undergraduate students from the University of York took part in this experiment (3 male, 27 female), receiving course credit or a payment of £2 for their participation. Participants were 18 – 21 years old (mean = 19.33; S.D. = 0.80) and were required to have lived in the UK for at least 18 years.

Design. The experimental variables comprised face type (with three levels; whole face, internal features or external features), AoA (with two levels; early or late) and gender of face (with two levels; male or female) and these were explored in a repeated measures design. The order of the face type conditions was counterbalanced across participants yielding six different predefined orders of presentation.

The dependent variables, response latency and accuracy, were the products of a familiarity decision task.

Materials. Sixty faces (30 early acquired, 30 late acquired) were selected from the dated database. Although, if a face had been acquired since 1996, images were taken from the current database to depict these items. None of the current images used in Experiments 5 and 6 were used in Experiments 7 and 8. At each level of AoA, there was an equal number of male and female faces. Independent t-tests confirmed that the four sets (early acquired male, early acquired female, late acquired male and late acquired female) were matched for frequency, distinctiveness and likeness, and the two early sets and two late sets were each matched for AoA (in each case $t \leq 1.022$). See Table 6.1 for summary data concerning the creation of experimental sets.

Faces in the internal and external conditions were edited in the same manner as those items featured in Experiments 5 and 6. Examples of the stimuli seen in each of the three conditions are shown on p 141.

The experimental stimuli were matched with an equal number of unfamiliar faces on gender, approximate age and racial group. The images were prepared in the same manner as the famous faces.

Table 6.1. *Mean Ratings as a Function of Experimental Set (Experiment 7).*

	Early acq. Male			Late acq. Male			Early acq. Female			Late acq. & female		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.85	(0.41)	2.14 - 3.41	3.94	(0.42)	3.57 - 4.95	2.73	(0.42)	2.26 - 3.43	4.05	(0.43)	3.48 - 4.76
Distinctiveness	2.91	(0.78)	1.70 - 4.81	2.89	(0.76)	1.88 - 4.30	2.81	(0.53)	2.00 - 3.63	2.76	(0.43)	2.08 - 3.45
Frequency	2.61	(0.44)	2.05 - 3.65	2.72	(0.62)	1.90 - 3.95	2.72	(0.77)	1.70 - 3.95	2.83	(0.68)	1.43 - 4.00
Likeness	3.61	(0.52)	2.24 - 4.46	3.69	(0.50)	2.88 - 4.42	3.56	(0.72)	2.46 - 4.69	3.61	(0.63)	2.36 - 4.52

Procedure and Apparatus. The familiarity decision task was executed in PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to the images presented in the centre of the screen. Each experimental trial began with a 500 ms blank screen prior to image onset. Faces were displayed until a response was made. 'Familiar' and 'unfamiliar' responses were logged using the 'z' and '/' keys on the keyboard, and keys pressed with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 pixels and the absolute dimensions of the whole face images were approximately 70 x 150 mm. Participants viewed these from a distance of approximately 30cm.

The 120 experimental trials (60 famous and 60 unfamiliar) were presented once as whole faces, once as internal features and once as external features. The face type conditions were blocked, and the order of the blocks was rotated yielding six orders of presentation: whole, internal, external (WIE); whole, external, internal (WEI); internal, whole, external (IWE); internal, external, whole (IEW); external, whole, internal (EWI); external, internal, whole (EIW). Each face type block was further subdivided into two blocks and the order of the two sub-blocks was randomized between participants; a measure to reduce the likelihood of associative priming.

Each block was preceded by instructions informing participants about the task requirements and the nature of the stimuli. For example, an excerpt from the internal condition instructions read "you are about to see a series of PART celebrity faces and unfamiliar faces presented one at a time. The parts will comprise internal features (eyes, nose, mouth). See example below". The instructions preceding the second and third blocks varied slightly from those read at the start as participants were additionally informed that the familiarity decision referred specifically to whether or not the face was famous. This line was inserted to counteract a growing sense of familiarity for the unfamiliar faces stemming from their repeated presentation over three blocks.

Forty practice items (20 famous and 20 unfamiliar) were administered prior to each of the three blocks. After the practice session, participants were reminded to respond "as quickly and accurately as possible". The experimental trials were directly preceded by 8 lead-in trials as a means of absorbing any initial hesitations. A representative sample of stimuli (early and late acquired; male and female faces) constituted practice and lead-in items to eliminate bias effects. None of these items were repeated as experimental items or included in the subsequent analysis. Practice and lead-in images were whole faces, internal features or external features as appropriate.

A familiarity verification exercise was completed before the experiment itself. Participants were presented with the current versions of the 120 faces featured in the experiment, in addition to their names, and asked to indicate at their own pace "which faces are familiar to you (i.e. you recognize them as being famous)". They were informed that the task was not timed and accuracy was paramount. Coupled with the experiment, the session lasted approximately 25 minutes.

6.1.3 Results

Ten participants[20] had to be replaced after failing to meet the criteria which required participants to: a) produce 75% correct responses to famous faces in the whole face condition and familiarity verification session; b) produce 75% correct responses to famous faces and unfamiliar faces pooled together, in the whole face condition and familiarity verification session; c) produce 75% correct responses to unfamiliar faces in the internal and external conditions (ensuring that despite the high error rates to famous faces in these conditions, correct 'familiar' responses are likely to be valid and not the product of a response strategy).

The error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions made to known faces (a face is 'known' if it is successfully recognized in the familiarity verification session); b) a failure to recognise a face in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect familiarity decisions made

[20] Six of these participants were 18 years old indicating that the images used may not have been optimal for this cohort of subjects: the face database had been created nearly 2 years prior to this experiment. Efforts were made to target participants over 18 years old in the subsequent experiments.

to known faces, 176 (9.8%) whole face responses, 444 (24.7%) internal feature responses, and 691 (38.4%) responses made to external features, were removed. In addition, 16 (0.9%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. Outliers slower than 3 standard deviations away from the mean and faster than 300ms were removed. As a consequence, 33 (1.8%) whole face responses, 22 (1.2%) internal feature responses and 28 (1.6%) external feature responses were removed. Outliers were calculated for each face type separately.

Only correct familiarity decisions and those made to faces successfully recognized in the familiarity verification session, were submitted to the latency analysis. Table 6.2 shows the mean correct response latencies in each condition of the experiment in addition to associated error rates. For inspection purposes, Table 6.2 additionally reports '% misclassifications' which denotes incorrect familiarity decisions made to known faces.

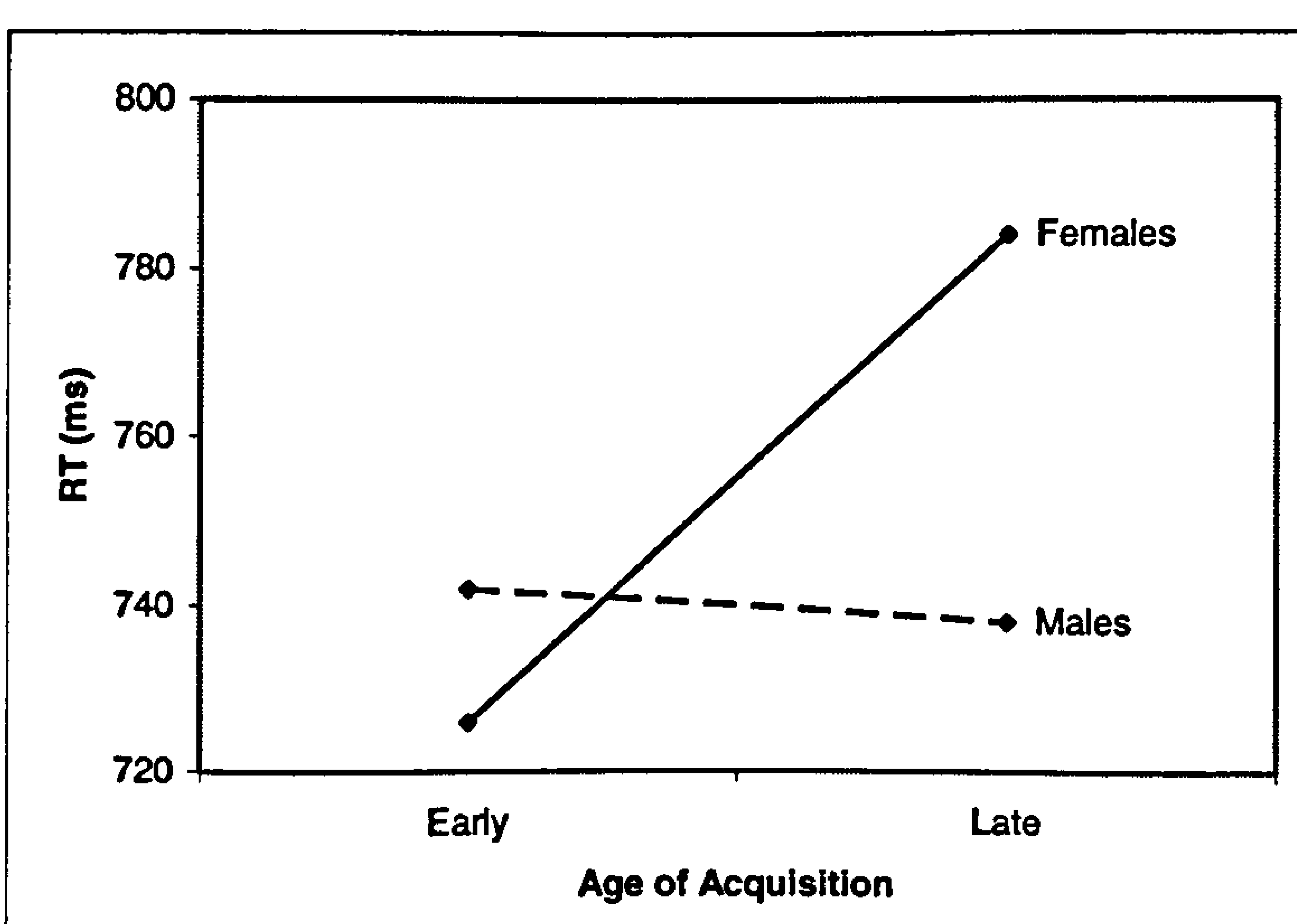
Table 6.2. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Experiment 7).*

		Early Acquired			Late acquired			Unfamiliar		
Stimulus		Male	Female	Mean	Male	Female	Mean	Male	Female	Mean
Whole	*Mean RT*	635	610	623	627	664	645	638	650	644
	SD	85	86	82	90	87	82	83	88	83
	% Error	13.1	7.3	10.2	11.6	18.0	14.8	6.6	7.9	7.2
	% Misclass.	9.7	6.2	7.9	8.6	15.0	11.8	3.5	3.7	3.6
Internal	*Mean RT*	788	800	793	773	816	794	777	805	791
	SD	128	130	116	120	131	117	133	126	128
	% Error	22.4	28.7	25.3	24.4	30.4	28.0	7.6	6.4	7.0
	% Misclass.	20.8	26.7	23.5	22.5	29.1	26.0	5.1	2.8	3.9
External	*Mean RT*	805	768	789	813	871	840	808	814	809
	SD	140	120	119	153	191	151	166	170	162
	% Error	41.8	34.7	38.2	43.1	43.1	43.1	14.4	11.1	12.8
	% Misclass.	39.5	33.6	36.6	41.7	40.0	40.8	11.4	7.7	9.5

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors, and by-items with face type, AoA and face gender as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of face type was significant in both analyses, $F_1(2,58) = 57.02$, MSe = 20,551.82, $p<.001$; $F_2(2,167) = 62.46$, MSe = 11,448.92, $p<.001$, confirming the observed tendency for wholes (634 ms) to be recognized more quickly than internals (794 ms), and internals more quickly than externals (814 ms). With alpha set at .05, Tukey tests revealed that whole faces were recognised significantly faster than internal and external features, but the difference between the latter face types was not significant (HSD = 89). The main effect of AoA was significant by-subjects, $F_1(1,29) = 14.36$, MSe = 4,314.36, $p = .001$, with early acquired faces (735 ms) recognized significantly faster than late acquired faces (761 ms). This was not significant by-items, $F_2(1,167) = 2.03$, MSe = 11,448.92, $p = .16$. The main effect of gender was not significant in either analysis, $F_1(1,29) = 2.12$, MSe = 8,910.58, $p = .16$; $F_2(1,167) = 0.01$, MSe = 11,448.92, $p = .92$.

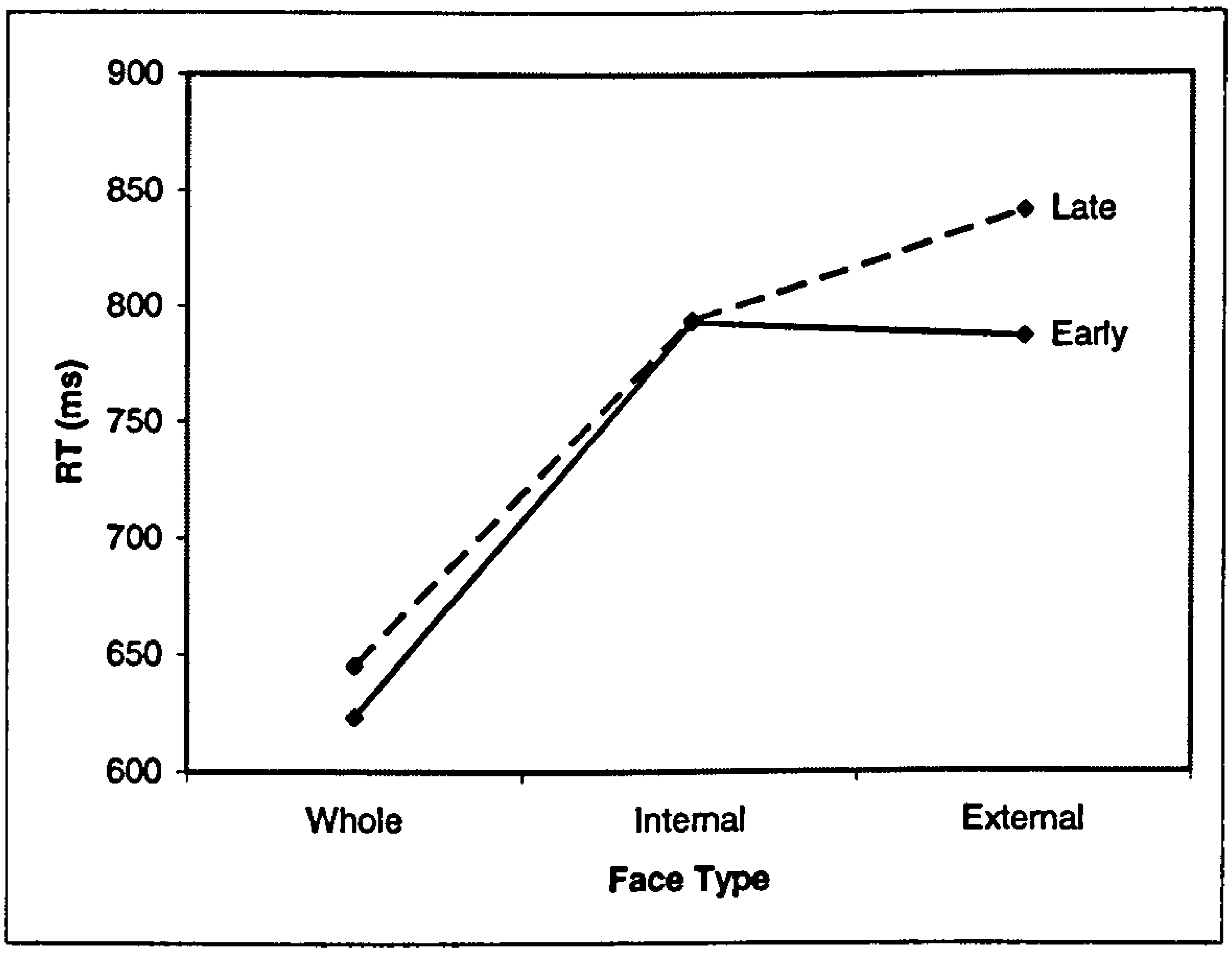
A significant two-way interaction between AoA and gender emerged from the by-subjects analysis, $F_1(1,29) = 18.51$, MSe = 4,749.42, $p<.001$, see Figure 6.1. *A posteriori* Tukey tests (HSD = 48) demonstrated a significant effect of AoA on female faces, but not on male faces. The effect of gender approached significance for late acquired faces, but did not modulate responses to early acquired faces.

Figure 6.1. *Effect of AoA on Familiarity Decision RTs as a Function of Gender (Experiment 7).*



The by-subjects analysis also yielded a significant two-way interaction between AoA and face type, $F_1(2,58) = 7.23$, MSe = 3,135.20, $p<.01$, see Figure 6.2. A simple main effects analysis demonstrated an effect of AoA on wholes, $F_1(1,29) = 12.99$, MSe = 588.20, $p = .001$, and externals, $F_1(1,29) = 13.43$, MSe = 3,432.95, $p = .001$, but internal feature performance was not affected by AoA, $F_1(1,29) = 0.01$, MSe = 1,275.92, $p = .93$. Additionally, early, $F_1(2,58) = 64.01$, MSe = 4,390.74, $p<.001$, and late acquired faces, $F_1(2,58) = 42.50$, MSe = 7,447.75, $p<.001$, were affected by face type. Tukey tests demonstrated that whole faces were recognised significantly faster than internal and external features, but the latter face types were not significantly different. This pattern emerged for the early (HSD = 41) and late acquired faces (HSD = 54). Neither analysis yielded any further interactions.

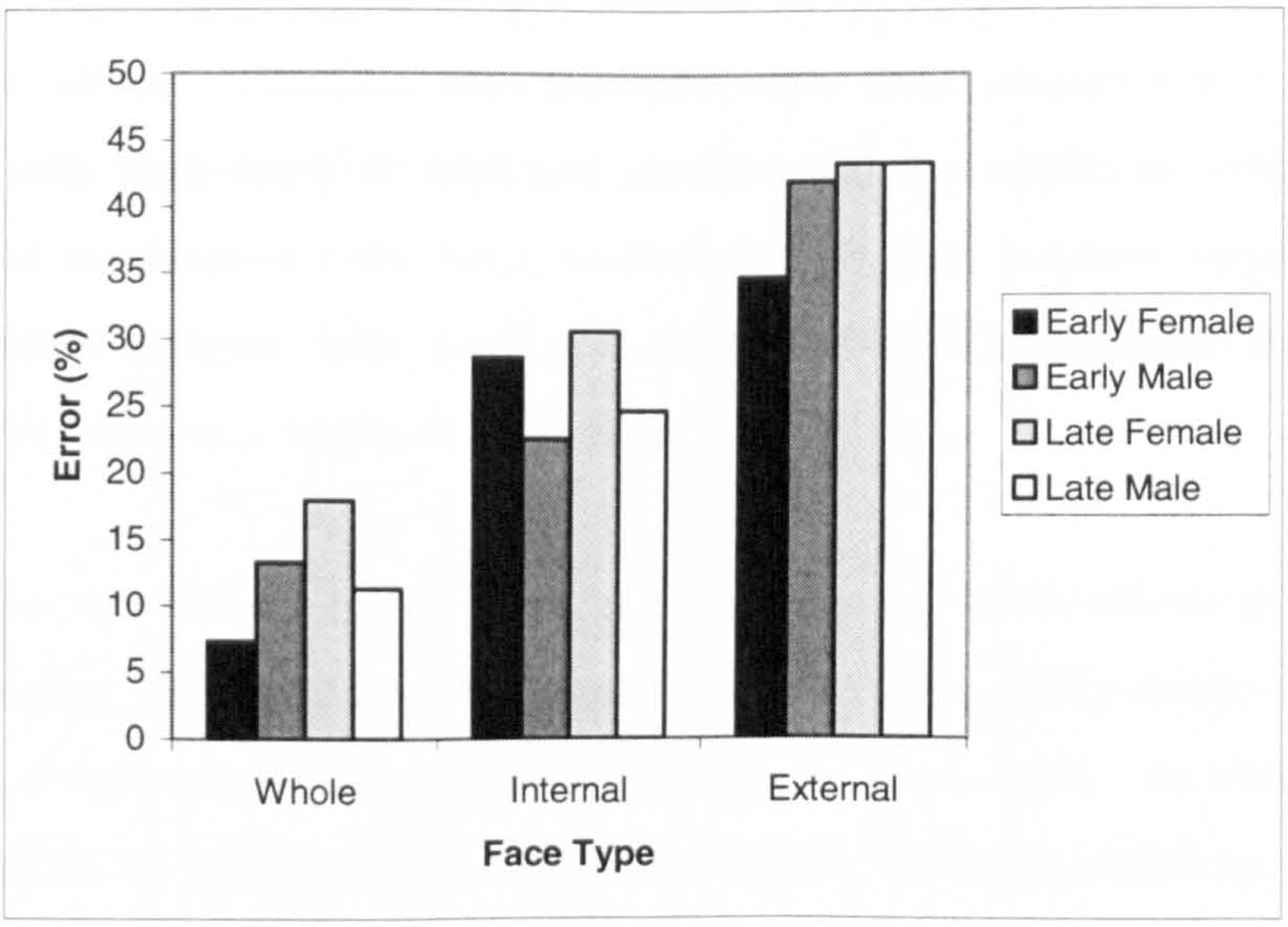
Figure 6.2. *Effect of AoA on Familiarity Decision RTs as a Function of Face Type (Experiment 7).*



Errors. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subject factors, and by-items with face type, AoA and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,58) = 54.56$, MSe = 9.82, $p<.001$; $F_2(2,168) = 28.37$, MSe = 37.76, $p<.001$, confirming the observed tendency for wholes (mean n = 1.9, 12.7%) to be recognized more accurately than internals (mean n = 4.0, 26.7%), and internals more accurately than externals (mean n = 6.1, 40.7%). Tukey tests (HSD = 1.9) revealed that these differences were significant. The main effect of AoA was significant by-subjects, $F_1(1,29) = 8.99$, MSe = 3.22, $p<.01$, with early acquired faces (mean n = 3.7, 24.7%) recognised more accurately than late acquired faces (mean n = 4.3, 28.7%). By-items, AoA was not significant, $F_2(1,168) = 1.14$, MSe = 37.76, $p = .29$. The main effect of gender failed to reach significance in either analysis, $F_1(1,29) = 0.34$, MSe = 5.49, $p = .56$; $F_2(1,168) = 0.24$, MSe = 37.76, $p = .63$.

Significant two-way interactions between AoA and gender, $F_1(1,29) = 7.71$, MSe = 2.67, $p = .01$, and face type and gender, $F_1(2,58) = 7.10$, MSe = 2.25, $p<.01$, emerged from the by-subjects analysis. These were incorporated into a three-way interaction between AoA, gender and face type which almost attained significance, $F_1(2,58) = 3.10$, MSe = 2.13, $p = .053$, see Figure 6.3. As a means of refining the extent of the AoA effect, the data from each face type were submitted independently to 2 x 2 within-subjects ANOVAs with AoA and gender as factors.

Figure 6.3. *Relationship Between AoA and Gender as a Function of Face Type, Indexed by Familiarity Decision Accuracy (Experiment 7).*



In the whole condition, whilst gender was not significant, $F_1(1,29) = 0.05$, MSe = 1.68, $p = .83$, AoA was significant, $F_1(1,29) = 7.56$, MSe = 1.85, $p = .01$, as was the interaction, $F_1(1,29) = 14.76$, MSe = 1.71, $p = .001$. Tukey tests (HSD = 0.8) revealed that female faces were affected by AoA, but males were not affected. In addition, early acquired females received fewer errors than early acquired males, and this gender effect reversed for late acquired faces. In the internals analysis, whilst gender was significant, $F_1(1,29) = 5.71$, MSe = 4.42, $p<.05$, AoA, $F_1(1,29) = 0.96$, MSe = 2.51, $p = .34$, and the interaction, $F_1(1,29) = 0.00$, MSe = 2.73, $p = .96$, were not significant. An analysis of the externals, revealed a main effect of AoA, $F_1(1,29) = 6.84$, MSe = 2.36, $p = .01$, but no significant main effect of gender, $F_1(1,29) = 2.19$, MSe = 3.90, $p = .15$. Further, the interaction was not significant, $F_1(1,29) = 3.44$, MSe = 2.48, $p = .07$. No interactions emerged from the by-items analysis.

6.1.3.1 Participant Gender Analysis

Only 10% of the participants tested were males, thus prohibiting a sensible analysis of participant gender.

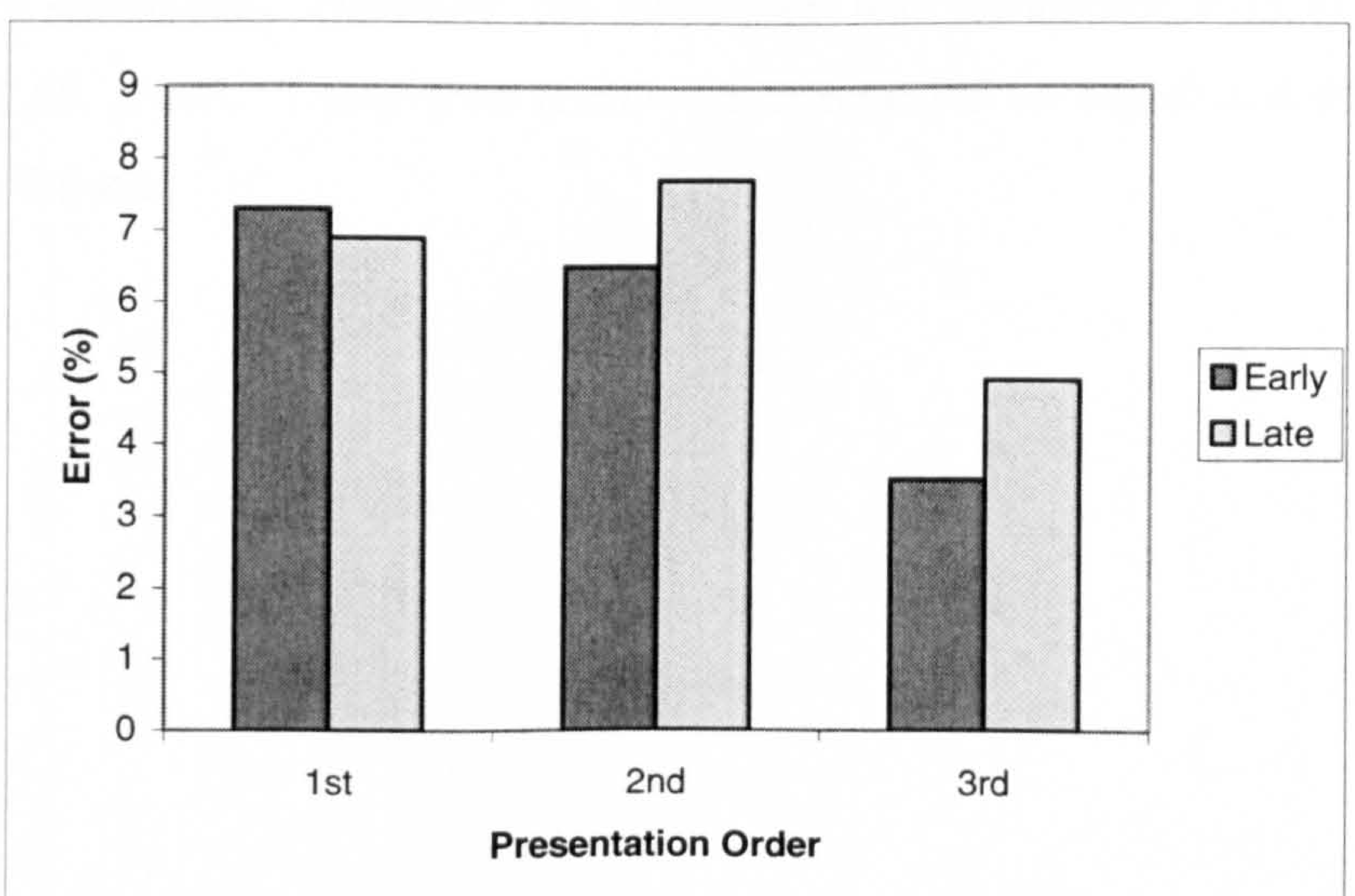
6.1.3.2 Presentation Order Analysis

A collection of subsidiary ANOVAs explored the effect of presentation order on each face type separately to check whether or not presentation order exerted systematic effects. Reaction time and error data were submitted to by-subjects analyses with AoA (early or late) and gender (male or female) as within-subject factors and presentation order (first, second or third) as a between-subjects factor, and by-items analyses with AoA and gender as between-subjects factors and presentation order as a within-subjects factor.

Reaction Times. Presentation order failed to modulate effects of gender in the whole face and internal feature conditions. However, the by-items analysis of externals recovered an interaction between gender and order. In short, gender effects failed to emerge during first, second or third presentations, however presentation order affected female and male face recognition latency: reaction times decreased significantly over successive presentations of female faces whereas male faces were only recognised significantly faster by the third presentation. Importantly, effects of AoA were not modulated by presentation order, regardless of face type.

Errors. Presentation order failed to modulate effects of AoA or gender in the internal condition. However, the by-subjects analysis of whole faces recovered a significant interaction between gender and order: gender effects failed to emerge during first, second or third presentations. Whilst there was no effect of presentation order on female faces, male faces were recognised increasingly accurately from first to second presentations but further increases in accuracy were not significant. In the external condition, AoA interacted significantly with order by-subjects and by-items, $F_1(2,27) = 4.79$, MSe = 1.87, $p<.05$; $F_2(2,112) = 5.13$, MSe = 1.21, $p<.01$, see Figure 6.4. A simple main effects analysis revealed that AoA effects were restricted to second, $F_1(1,27) = 7.08$, MSe = 0.93, $p = .01$, and third presentations, $F_1(1,27) = 10.49$, MSe = 0.93, $p<.01$, failing to reach significance during the first presentation, $F_1(1,27) = 0.66$, MSe = 0.93, $p = .43$. In addition, presentation order affected early acquired faces, $F_1(2,27) = 7.93$, MSe = 40.51, $p<.01$, but failed to modulate late acquired face recognition accuracy, $F_1(1,27) = 2.82$, MSe = 7.46, $p = .08$. Tukey tests (HSD = 7.1) demonstrated that there were no significant increases in early acquired face recognition accuracy with successive presentations.

Figure 6.4. *Effect of AoA on Familiarity Decision Accuracy as a Function of Presentation Order (Experiment 7).*



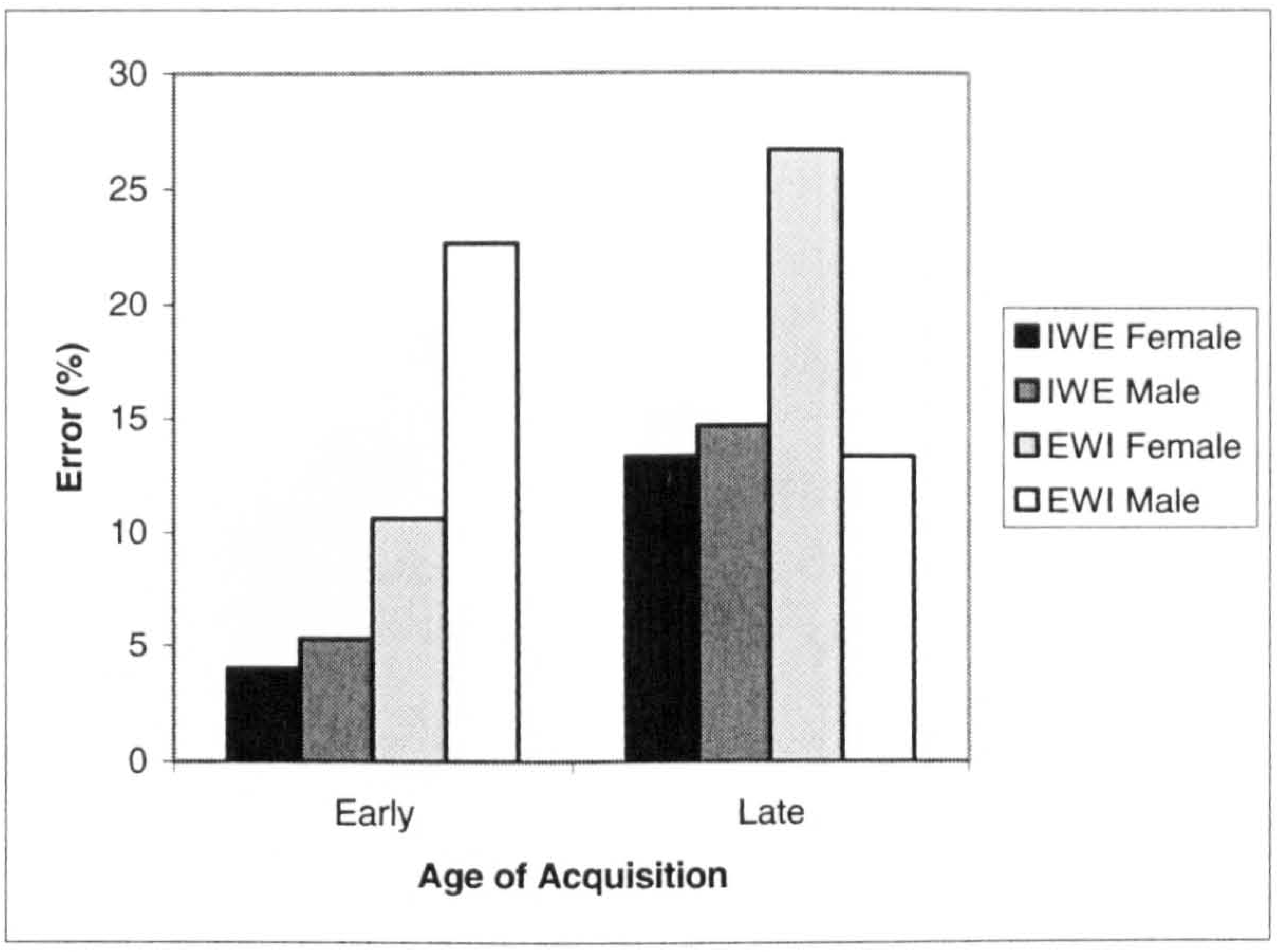
6.1.3.3 Analysis of Second Presentation Groups

The variable 'presentation order' was probed further to investigate the impact of being preceded by one face type compared with another. Block 2 data were subject to scrutiny. Reaction time and error data for each face type were independently submitted to mixed-design ANOVAs. In the whole face condition, for example, data were analysed by-subjects with AoA (early or late) and gender (male or female) as within-subject factors and group (IWE or EWI) as a between-subjects factor.

Reaction Times. Group failed to modulate effects of AoA or gender.

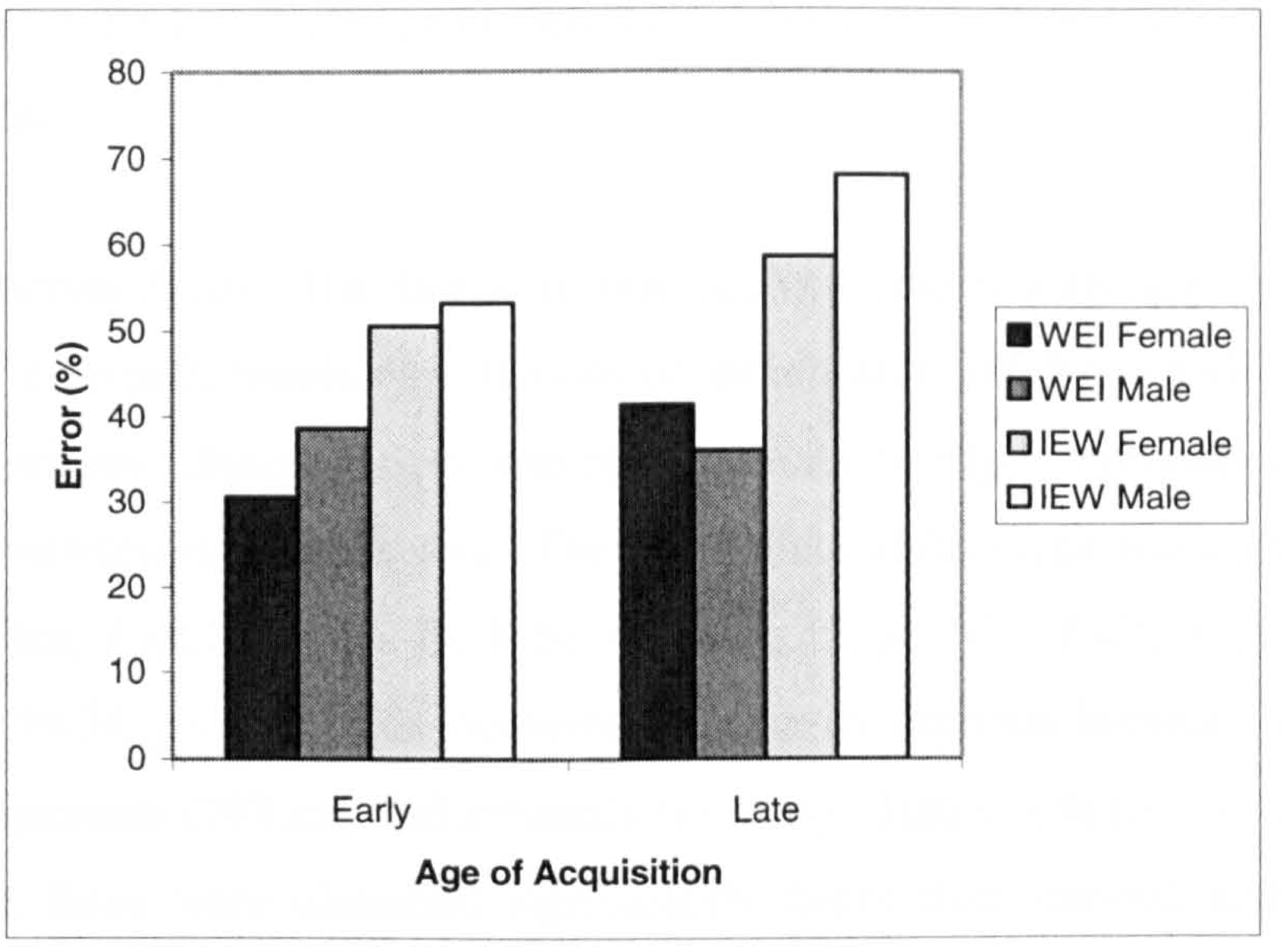
Errors. Group failed to modulate effects of AoA or gender in the internal condition. However, a significant interaction between group, AoA and gender emerged in the whole face analysis, $F_1(1,8) = 8.81$, MSe = 1.03, $p<.05$, see Figure 6.5. To evaluate systematic effects of group, data from each group were submitted independently to 2 x 2 within-subjects ANOVAs with AoA and gender as factors. In the IWE group, AoA was significant, $F_1(1,4) = 12.25$, MSe = 0.80, $p<.05$, but gender failed to attain significance, $F_1(1,4) = 0.17$, MSe = 0.20, $p = .70$, as did the interaction, $F_1(1,4) = 0.00$, MSe = 0.50, $p = 1.00$. In the EWI group, neither AoA, $F_1(1,4) = 0.31$, MSe = 4.00, $p = .61$, nor gender, $F_1(1,4) = 0.02$, MSe = 2.30, $p = .89$, were significant. However, the interaction was significant, $F_1(1,4) = 11.65$, MSe = 1.55, $p<.05$. Tukey tests (HSD = 3.2) revealed no significant differences between the sets.

Figure 6.5. *Relationship Between AoA and Gender as a Function of Group, Indexed by Familiarity Decision Accuracy (Experiment 7).*



In the external features analysis, an interaction between group, AoA and gender approached significance, $F_1(1,8) = 5.00$, MSe = 1.13, $p<.056$, see Figure 6.6. Data from each group were submitted independently to 2 x 2 within-subjects ANOVAs with AoA and gender as factors. In the WEI group, neither AoA, $F_1(1,4) = 0.53$, MSe = 3.43, $p = .51$, nor gender, $F_1(1,4) = 0.04$, MSe = 5.08, $p = .55$, reached significance, and neither did the interaction, $F_1(1,4) = 5.71$, MSe = 0.88, $p = .08$. In the IEW group, AoA was significant, $F_1(1,4) = 44.46$, MSe = 0.33, $p<.01$. However, gender, $F_1(1,4) = 1.67$, MSe = 2.43, $p = .27$, and the interaction, $F_1(1,4) = 0.91$, MSe = 1.38, $p = .39$, were not significant.

Figure 6.6. *Relationship Between AoA and Gender as a Function of Group, Indexed by Familiarity Decision Accuracy (Experiment 7).*



6.1.3.4 Familiarity Analysis

Post-hoc analyses were carried out on the famous face and unfamiliar face data to assess the impact of familiarity on familiarity decisions. These analyses were carried out to provide a comparison with post-hoc familiarity analyses of gender classification performance (Experiment 8).

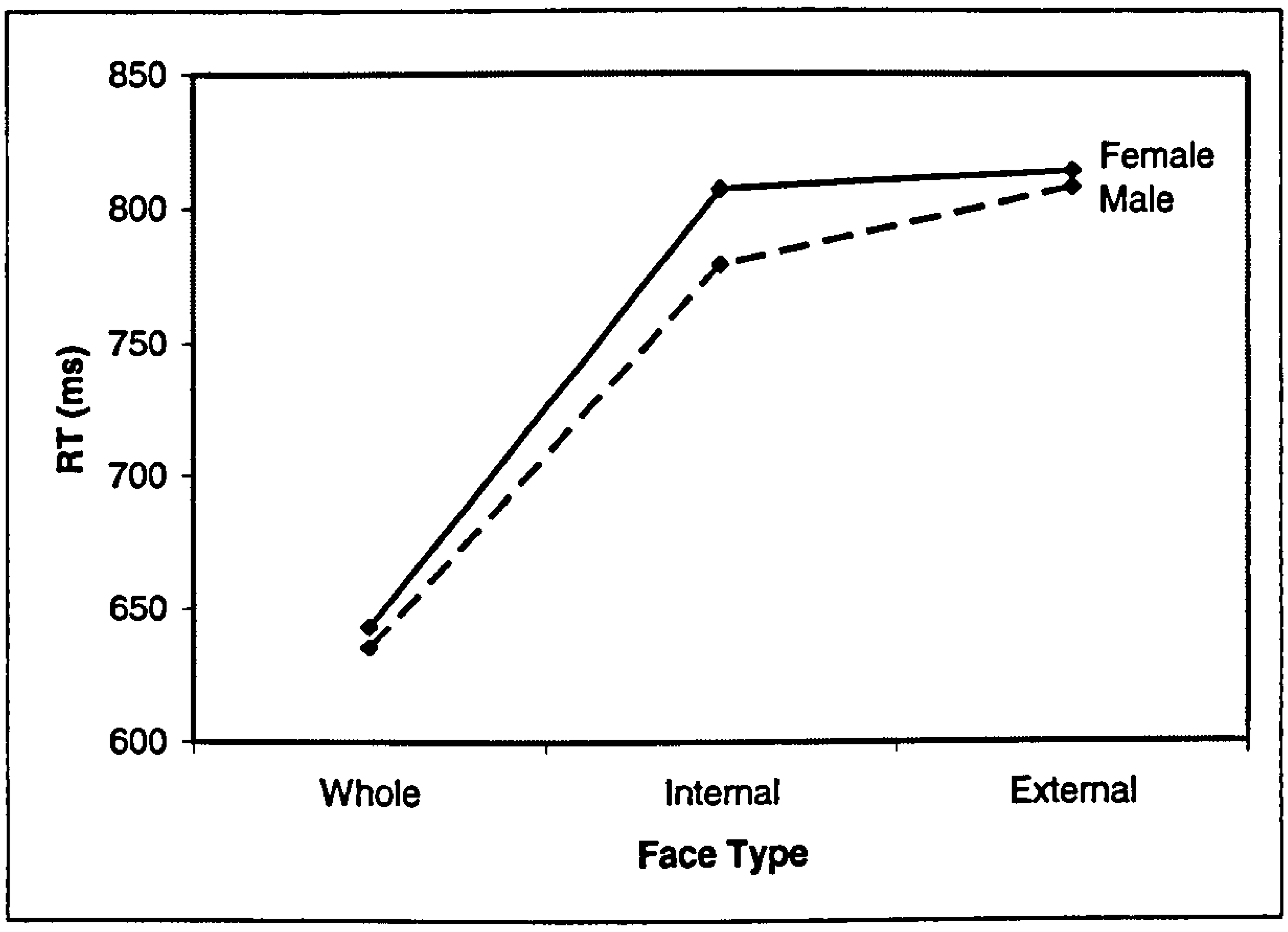
For the unfamiliar faces, the error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions made to faces correctly identified as 'unfamiliar' in the familiarity verification session; b) a failure to correctly identify a face as 'unfamiliar' in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect familiarity decisions, 64 (3.6%) whole face responses, 70 (3.9%) internal feature responses, and 169 (9.4%) external feature responses, were removed. In addition, 30 (1.7%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. Further, 36 (2.0%) responses from the whole condition were deemed outliers, as were 25 (1.4%) internal condition responses and 30 (1.7%) external.

Only correct familiarity decisions, and those made to faces correctly classified as 'unfamiliar' in the familiarity verification session, were submitted to the latency analysis. See Table 6.2 for mean famous and unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous or unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,58) = 95.40$, MSe = 11,270.11, $p<.001$; $F_2(2,347) = 164.91$, MSe = 7,199.34, $p<.001$, with increasing familiarity decision latencies to wholes (639 ms), internals (793 ms) and externals (811 ms). Tukey tests ($\alpha = .05$) revealed that whole faces were classified significantly faster than internal and external features, but the difference between the latter face types was not significant (HSD = 66). Familiarity was not significant by-subject or by-items, $F_1(1,29) = 0.01$, MSe = 61,515.09, $p = .94$; $F_2(1,347) = 2.99$, MSe = 7,199.34, $p = .08$. The main effect of gender was significant by-subjects, $F_1(1,29) = 6.57$, MSe = 2,674.82, $p<.05$, with male faces (741 ms) classified faster than female faces (755 ms). However, the effect was not replicated by-items, $F_2(1,347) = 1.18$, MSe = 7,199.34, $p = .28$.

A two-way interaction between face type and gender approached significance in the by-subjects analysis, $F_1(2,58) = 3.11$, MSe = 1,479.03, $p = .052$, see Figure 6.7. A simple main effects analysis revealed no effect of gender on whole faces, $F_1(1,29) = 3.12$, MSe = 313.17, $p = .09$, or externals, $F_1(1,29) = 0.32$, MSe = 1,482.71, $p = .58$. However, gender modulated responses to internals, $F_1(1,29) = 11.66$, MSe = 1,023.36, $p<.01$. Further, effects of face type were shown to be significant for female, $F_1(2,58) = 74.18$, MSe = 3,789.83, $p<.001$, and male faces, $F_1(2,58) = 99.98$, MSe = 2,583.67, $p<.001$. In the case of both female (HSD = 38) and male faces (HSD = 32) Tukey tests attributed this significance to a difference between wholes and the other face types with the difference between internals and externals not reaching significance.

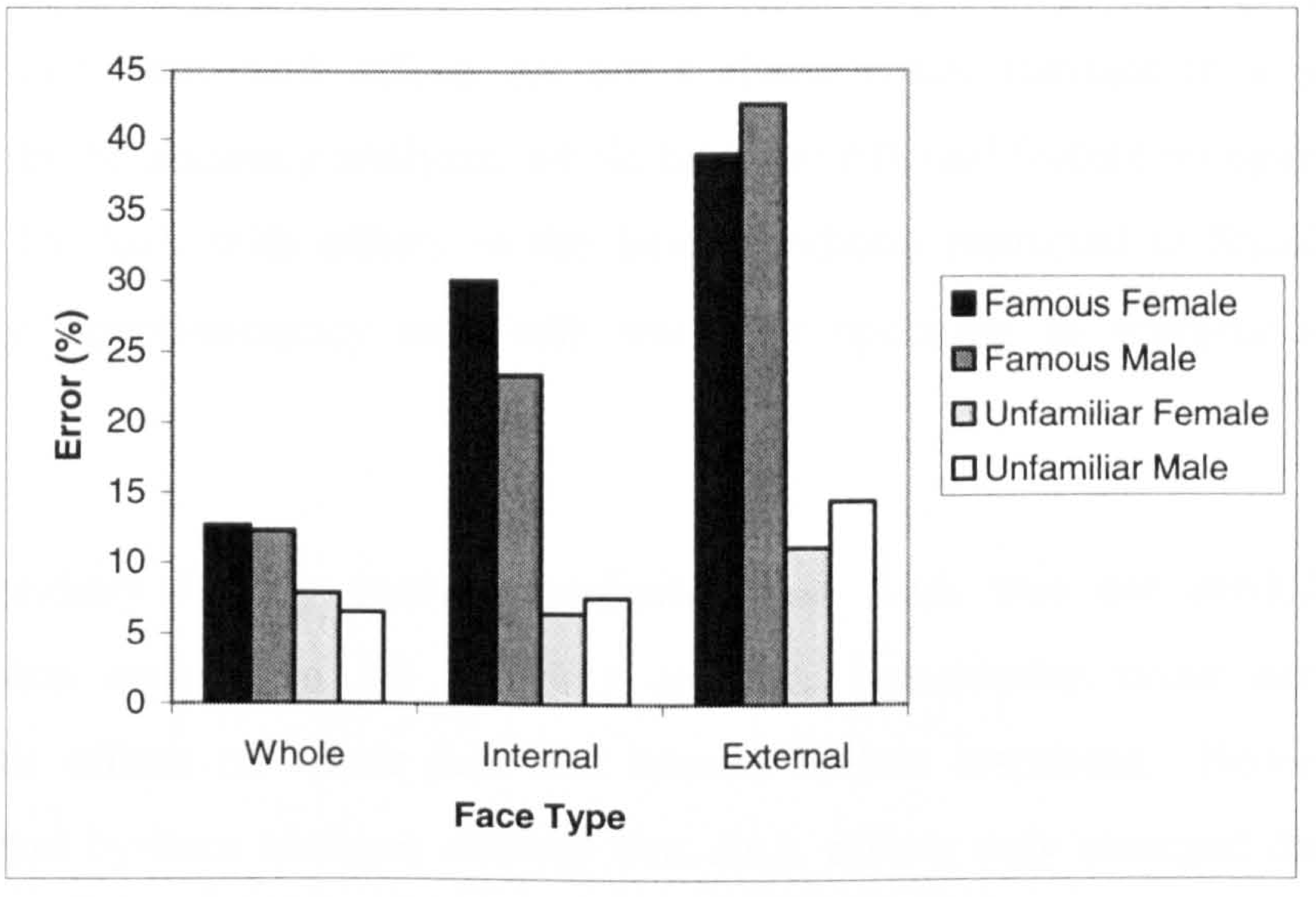
Figure 6.7. *Effect of Gender on Familiarity Decision RTs as a Function of Face Type (Experiment 7).*



Errors. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous or unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,58) = 60.75$, MSe = 12.80, $p<.001$; $F_2(2,348) = 36.71$, MSe = 21.13, $p<.001$, with increasing errors to wholes (mean n = 3.0, 10.0%), internals (mean n = 5.1, 17.0%) and externals (mean n = 8.0, 26.7%). Tukey tests (HSD = 2.2) revealed that external features were recognised significantly less accurately than whole faces and internal features, but the difference between the latter face types was not significant. The main effect of familiarity was significant in both analyses, $F_1(1,29) = 62.37$, MSe = 40.54, $p<.001$; $F_2(1,348) = 111.74$, MSe = 21.13, $p<.001$, with unfamiliar faces (mean n = 2.7, 9.0%) classified more accurately than famous faces (mean n = 8.0, 26.7%). The main effect of gender was not significant in either analysis, $F_1(1,29) = 0.00$, MSe = 8.45, $p = .97$; $F_2(1,348) = 0.02$, MSe = 21.13, $p = .89$.

A two-way interaction between familiarity and face type was significant in both analyses, $F_1(2,58) = 34.39$, MSe = 10.35, $p<.001$; $F_2(2,348) = 16.71$, MSe = 21.13, $p<.001$. In the by-subject analysis alone, the interaction between face type and gender also reached significance, $F_1(2,58) = 12.28$, MSe = 2.27, $p<.001$. These by-subjects interactions were subsumed in a significant three-way interaction between face type, familiarity and gender, $F_1(2,58) = 4.30$, MSe = 3.75, $p<.05$, see Figure 6.8. As a means of refining the extent of the AoA effect, data from each face type were submitted independently to 2 x 2 within-subjects ANOVAs with familiarity and gender as factors.

Figure 6.8. *Relationship Between Face Type and Gender as a Function of Familiarity, Indexed by Familiarity Decision Accuracy (Experiment 7).*



In the whole condition, familiarity was significant, $F_1(1,29) = 7.79$, MSe = 9.66, $p<.01$. Gender was not significant, $F_1(1,29) = 0.86$, MSe = 2.19, $p = .36$, and neither was the interaction, $F_1(1,29) = 0.29$, MSe = 2.37, $p = .60$. In the internal condition, familiarity, $F_1(1,29) = 52.55$, MSe = 20.10, $p<.001$, and gender were significant, $F_1(1,29) = 4.22$, MSe = 4.94, $p<.05$. The interaction was also significant, $F_1(1,29) = 7.44$, MSe = 5.49, $p = .01$. Tukey tests (HSD = 1.7) demonstrated that effects of gender were restricted to famous faces with famous females attracting more errors than famous males. Further, familiarity affected male and females alike with more errors made to famous faces compared with unfamiliar faces. In the externals analysis, effects of familiarity, $F_1(1,29) = 66.98$, MSe = 31.48, $p<.001$, and gender, $F_1(1,29) = 5.64$, MSe = 5.87, $p<.05$, were significant. However, the interaction was not, $F_1(1,29) = 0.01$, MSe = 7.01, $p = .92$.

6.1.4 Discussion

Contrary to predictions, AoA effects were restricted to whole and external conditions, failing to modulate responses to internals. This is rather unfortunate and may compromise the translation of AoA effects onto gender decisions made to internals. In addition, AoA effects were restricted to female faces. Given that male faces were affected by AoA in the original familiarity decision task (Experiment 5), it is clear that AoA effects are not a characteristic intrinsic to a particular gender. In the accuracy analysis, whole face and external feature recognition was affected by AoA with effects in the latter condition restricted to female faces. Evidently, speed-accuracy trade-offs were not operating to compromise AoA effects.

The subsidiary latency analyses indicated that AoA was not modulated by presentation order. In the accuracy analysis, presentation order exerted no systematic effects on whole face and internal feature responses. However, by-subject and by-item analyses showed that AoA effects only emerged during the second and third presentations of external features. Likewise, there was an effect of being preceded by either one face type of another on the accuracy analysis of the whole and external conditions. When wholes had been preceded by internals, AoA affected accuracy. However, when wholes had been preceded by externals, AoA did not affect accuracy. Similarly, when externals had been preceded by internals, AoA affected accuracy but when externals were preceded by wholes, AoA effects did not ensue. Thus, AoA effects in the accuracy analysis appeared not to be particularly robust. The unreliable AoA effects throughout Experiment 7 were far from the ubiquitous effects predicted to modulate familiarity decisions. Moreover, this does not set a particularly strong precedence for the gender classification task.

Contrary to predictions and effects documented in the literature (Bruce et al., 1987a), familiarity did not affect reaction time performance. As predicted, unfamiliar faces were classified more accurately than familiar faces for all face types, reflecting response strategies rather than intrinsic properties of the stimuli.

6.2 Experiment 8

6.2.1 Introduction

Experiment 8 replicated Experiment 7 with gender decisions made to early and late acquired whole faces, internal features and external features. An effect of AoA on internal features was hypothesised and it was anticipated that these effects would be consistent across male and female faces. Having said this, the unreliable AoA effects found in Experiment 7 did not create a particularly robust baseline of AoA effects. To be consistent with Experiment 7, the familiarity verification session preceded the experiment proper.

In addition, an effect of familiarity was predicted to emerge from the post-hoc analysis, with famous faces classified according to gender faster and more accurately than unfamiliar faces. More specifically, this effect was predicted to emerge with internal features only.

6.2.2 Method

Participants. Thirty undergraduate students from the University of York took part in this experiment (10 male, 20 female), receiving course credit or a payment of £2 for their participation. Participants were 18 – 21 years old (mean = 19.83; S.D. = 0.75) and were required to have lived in the UK for at least 18 years.

Design. The experimental variables comprised face type (with three levels; whole face, internal features or external features), AoA (with two levels; early or late) and gender of face (with two levels; male or female) which were explored in a repeated measures design. The order of the face type conditions was counterbalanced across participants yielding six different predefined orders of presentation.

The dependent variables, response latency and accuracy were the products of a gender classification task.

Materials. See Experiment 7 details, as the same stimuli were employed in this gender classification task, p. 174.

Procedure and Apparatus. See Experiment 7 details, p. 176. Note, however, that during the experiment proper participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is male or female".

In addition, a familiarity decision task with the 120 intact experimental stimuli was administered after the experiment proper. This task was carried out to ensure that only participants who could recognise a sufficient number of dated images were included in the experiment. In Experiment 6, familiarity verification recognition performance provided a good indication of how many faces in the gender classification task were actually known by participants, and therefore which participants should be excluded. However, because Experiment 8 uses current images in the familiarity verification session and dated images in the experiment proper, familiarity verification performance does not provide a good measure of faces recognised during the gender classification task. Thus, it cannot be used to gauge participant eligibility. The entire session lasted approximately 25 minutes.

6.2.3 Results

With exception to 1 participant who had to be replaced, all participants met the criteria which required them to: a) produce 75% correct responses to famous faces in the whole face condition and speeded familiarity decision task; b) produce 75% correct responses to famous faces and unfamiliar faces pooled together, in the whole face condition and familiarity decision task; c) produce 75% correct responses to unfamiliar faces in the internal and external conditions.

The error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications made to known faces (a face is 'known' if it is successfully recognized in the familiarity verification session); b) a failure to recognise a face in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect gender classifications made to known faces, 54 (3.0%) whole face responses, 65 (3.6%) internal feature responses, and 90 (5.0%) responses made to external features, were removed. In addition, 21 (1.2%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. As a result of eliminating outliers slower than 3 standard deviations away from the mean and faster than 300ms, 22 (1.2%) whole face responses, 28 (1.6%) internal feature responses and 34 (1.9%) external feature responses were removed. Outliers were calculated for each face type separately.

Only correct gender classifications, and those made to faces successfully recognized in the familiarity verification session, were submitted to the latency analysis. Table 6.3 shows the mean correct response latencies in each condition of the experiment in addition to associated error rates. For inspection purposes, Table 6.3 additionally reports '% misclassifications' which denotes incorrect gender classifications made to faces classified correctly during the familiarity verification session.

Table 6.3. *Mean RTs and Percentage Errors for Gender Decisions in Each Condition (Experiment 8).*

		Early Acquired			Late acquired			Unfamiliar		
Stimulus		Male	Female	Mean	Male	Female	Mean	Male	Female	Mean
Whole	*Mean RT*	613	572	593	570	589	580	567	565	566
	SD	105	112	104	102	129	107	100	85	88
	% Error	7.6	4.4	5.9	3.8	6.0	4.9	7.7	11.3	9.5
	% Misclass.	4.7	1.8	3.2	2.0	3.9	2.9	1.5	4.1	2.8
Internal	*Mean RT*	639	664	652	656	660	659	641	665	653
	SD	116	121	113	127	117	114	108	136	118
	% Error	4.9	9.8	7.3	5.6	5.1	5.3	9.0	11.3	10.2
	% Misclass.	2.9	5.9	4.4	2.7	3.2	2.9	3.7	5.2	4.4
External	*Mean RT*	635	618	627	628	616	622	614	608	612
	SD	102	107	92	95	103	88	93	91	84
	% Error	6.2	6.2	6.2	6.2	13.6	9.9	7.3	13.4	10.4
	% Misclass.	3.8	3.1	3.5	2.9	10.3	6.6	1.7	6.1	3.9

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors, and by-items with face type, AoA and face gender as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of face type was significant in both analyses, $F_1(2,58) = 16.91$, MSe = 8,405.30, $p<.001$; $F_2(2,168) = 13.02$, MSe = 5,787.77, $p<.001$, with wholes (586 ms) classified according to gender more quickly than externals (624 ms), and externals more quickly than internals (655 ms). With alpha set at .05, Tukey tests revealed that whole faces were classified significantly faster than internal and external features, but the latter face types did not differ significantly (HSD = 57). The main effect of AoA was not significant in either analysis, $F_1(1,29) = 0.42$, MSe = 2,978.72, $p = .52$; $F_2(1,168) = 0.51$, MSe = 5,787.77, $p = .48$. Also, the main effect of gender was not significant in either analysis, $F_1(1,29) = 0.29$, MSe = 4,306.47, $p = .59$; $F_2(1,168) = 0.18$, MSe = 5,787.77, $p = .68$.

A significant three-way interaction between AoA, gender and face type emerged from the by-subjects analysis, $F_1(2,58) = 4.17$, MSe = 3,192.68, $p<.05$, see Figures 6.9 to 6.11. Data from each face type condition were independently scrutinized in 2 x 2 within-subject ANOVAs to probe the effect of AoA on gender.

Figure 6.9. *Effect of AoA on Gender Classification RTs to Whole Faces as a Function of Face Gender (Experiment 8).*

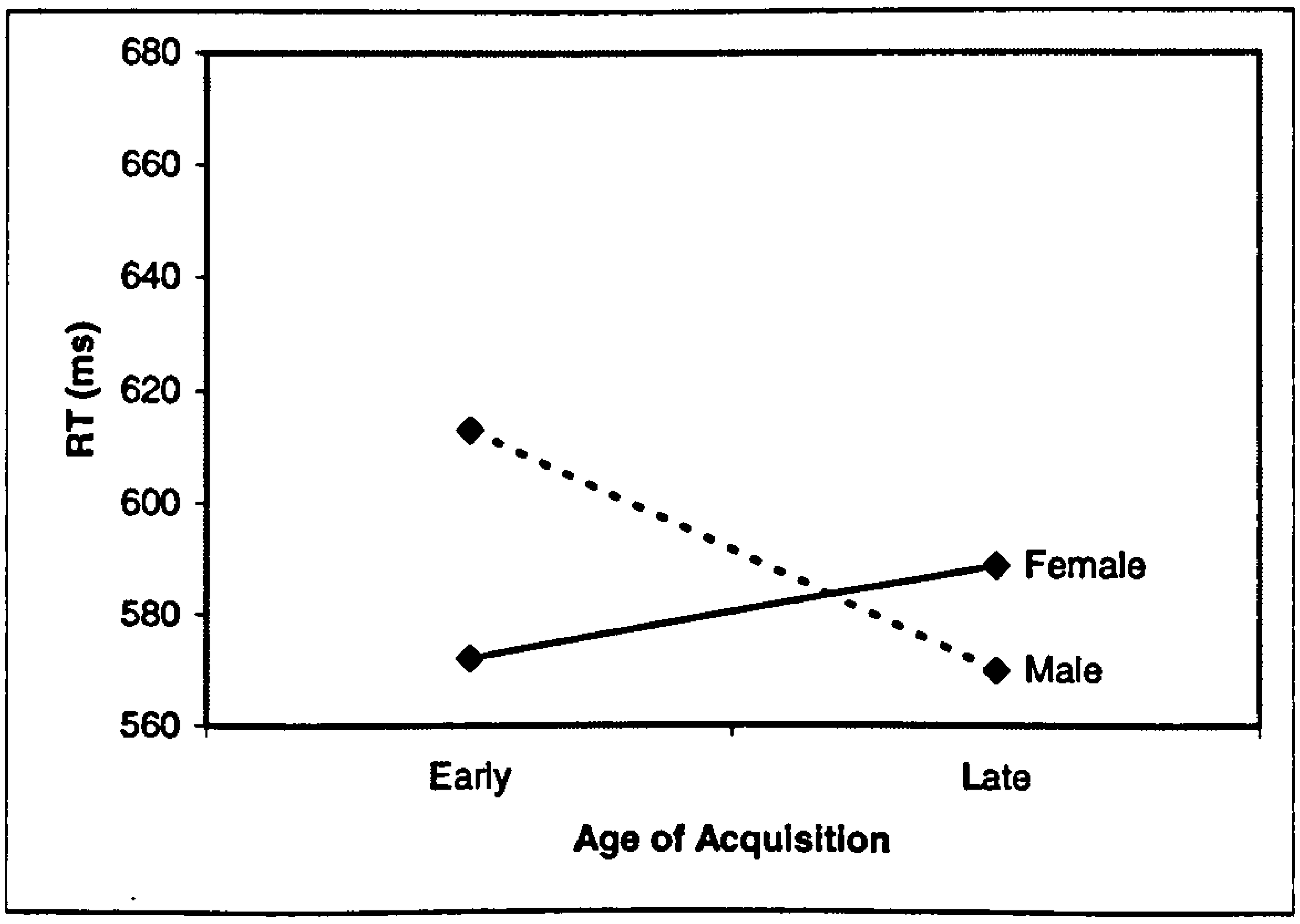


Figure 6.10. *Effect of AoA on Gender Classification RTs to Internal Features as a Function of Face Gender (Experiment 8).*

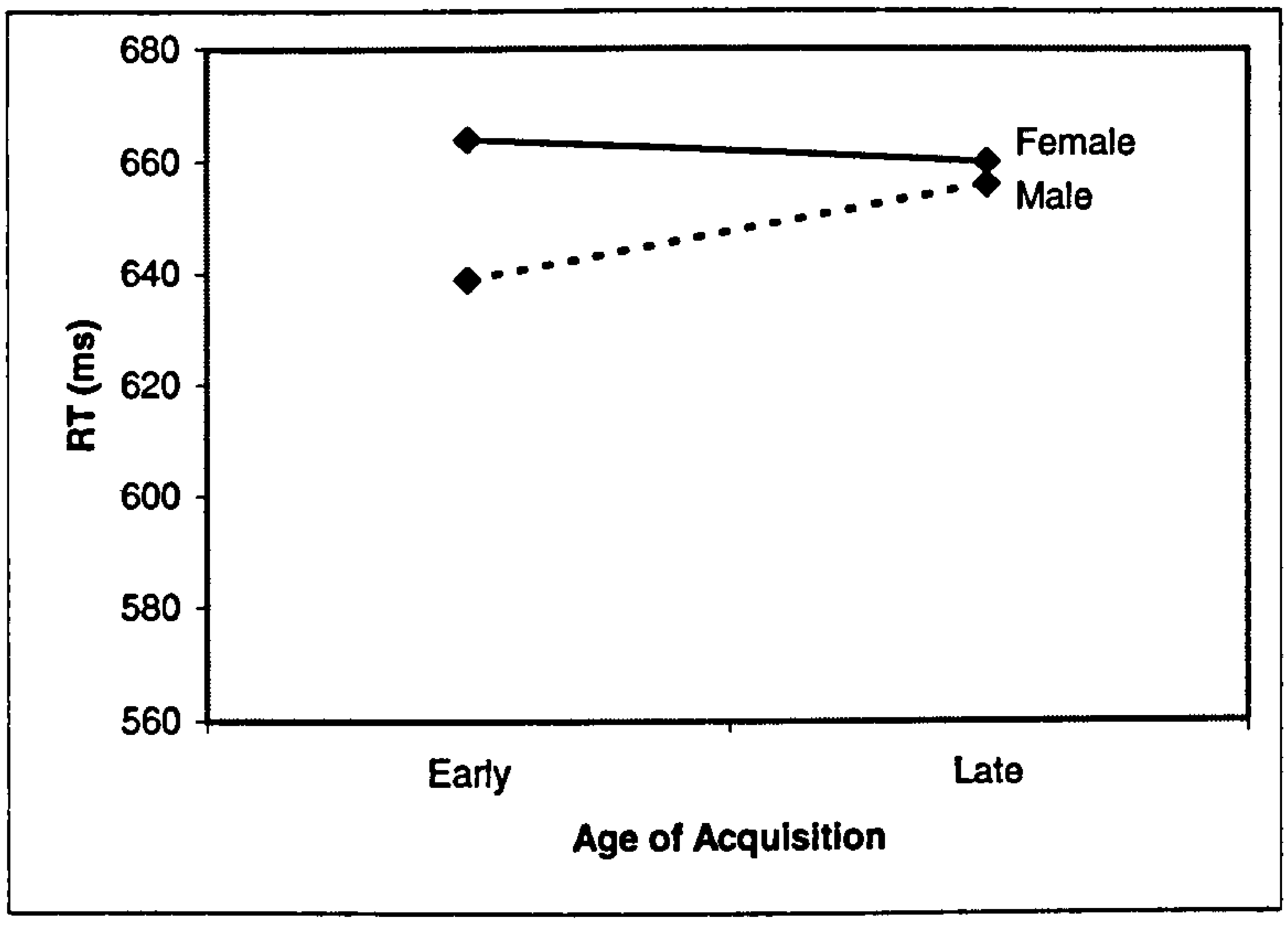
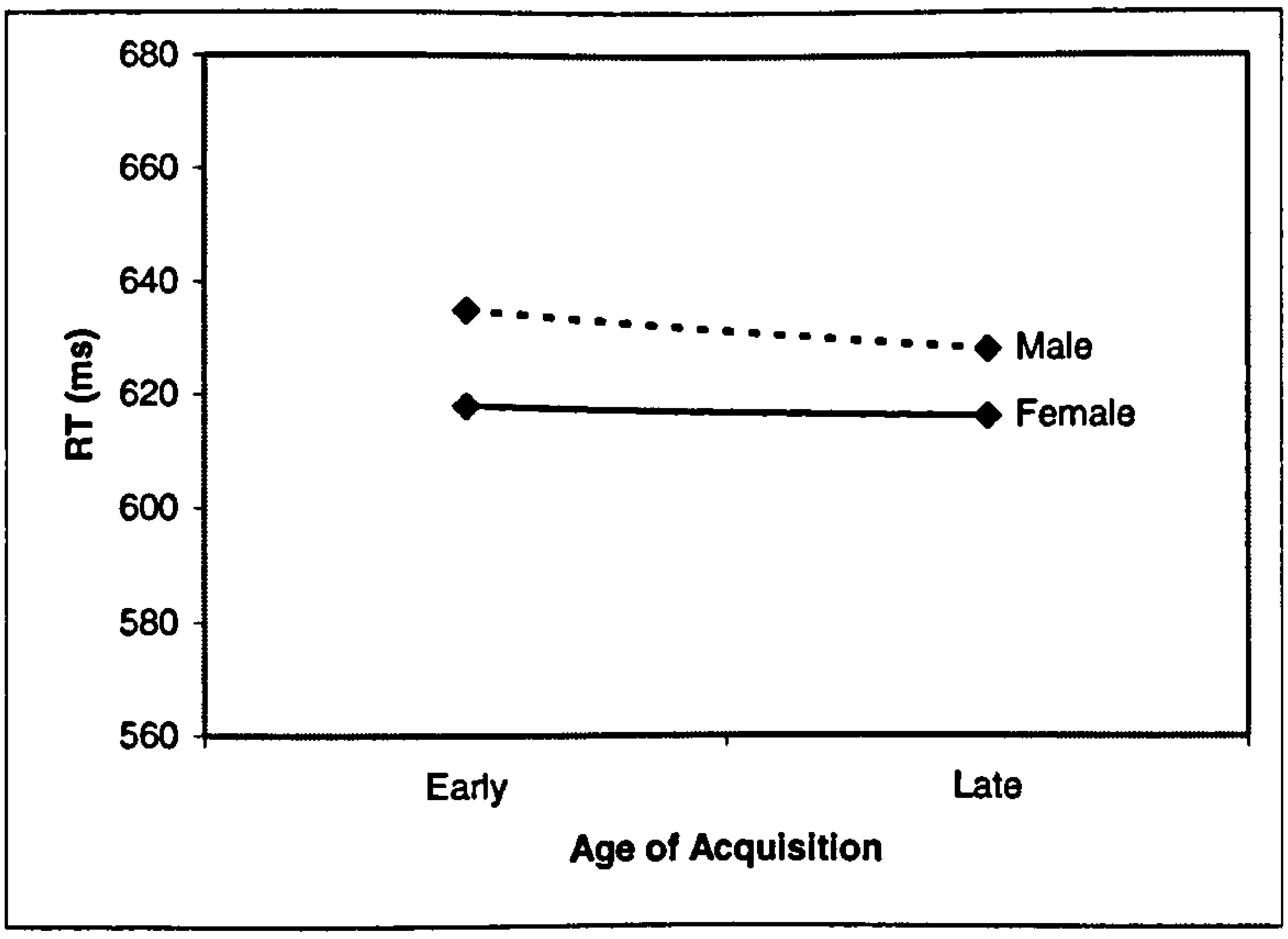


Figure 6.11. *Effect of AoA on Gender Classification RTs to External Features as a Function of Face Gender (Experiment 8).*

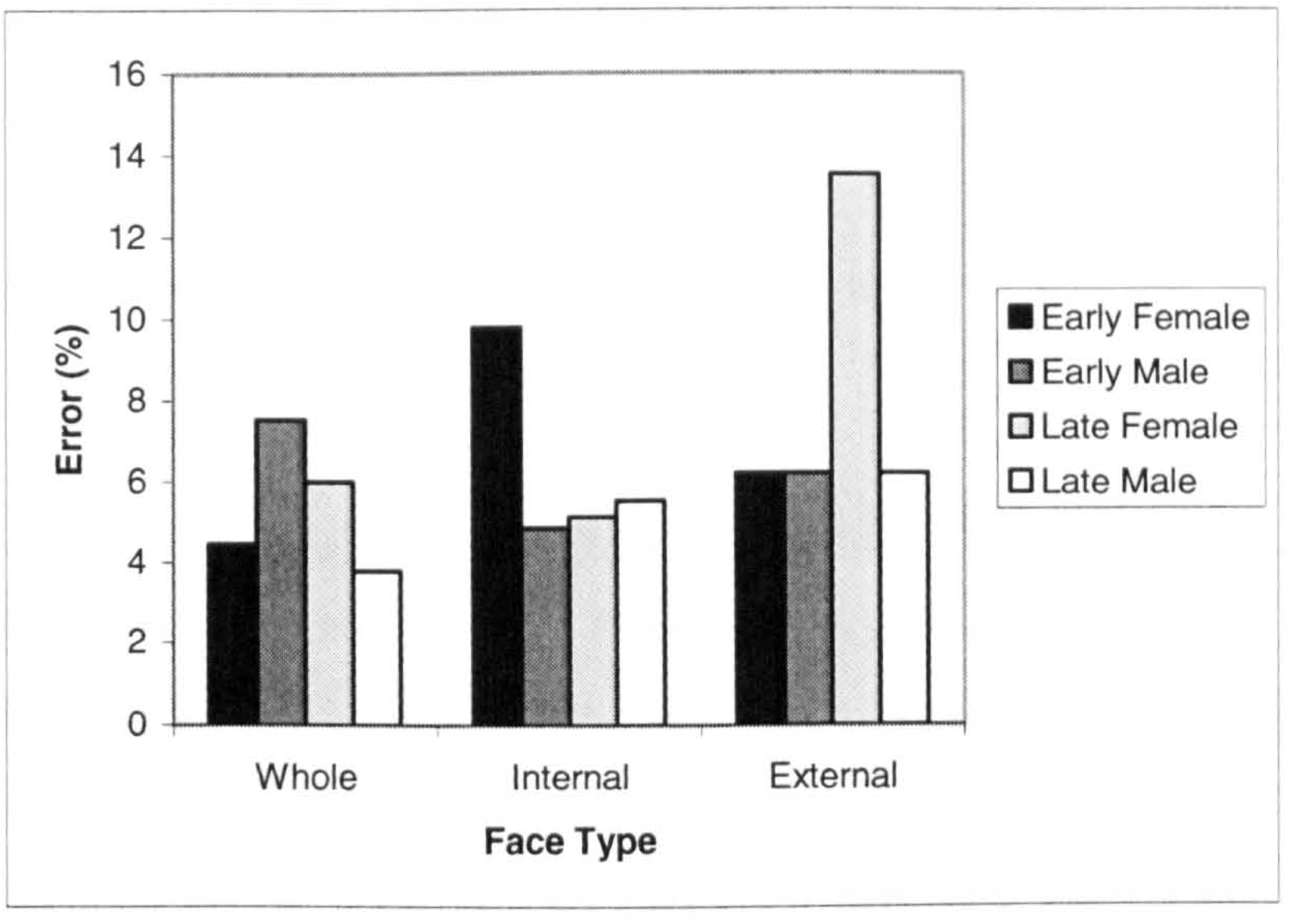


In the whole face condition, neither AoA, $F_1(1,29) = 2.92$, MSe = 1,887.64, $p = .10$, nor gender, $F_1(1,29) = 0.78$, MSe = 4,403.70, $p = .38$, were significant. However, the interaction was significant, $F_1(1,29) = 13.95$, MSe = 1,985.74, $p = .001$. Tukey tests (HSD = 31) showed that the effect of gender was restricted to early acquired faces, with female faces classified significantly faster than male faces, and that the AoA effect, although reversed, affected male faces only. In the analysis of internal features, neither AoA, $F_1(1,29) = 0.33$, MSe = 3,922.06, $p = .57$, nor gender, $F_1(1,29) = 1.82$, MSe = 3,391.49, $p = .19$, were significant. Also, the interaction was not significant, $F_1(1,29) = 1.00$, MSe = 3,558.41, $p = .33$. In the external features analysis, neither AoA, $F_1(1,29) = 0.18$, MSe = 2,964.88, $p = .67$, nor gender, $F_1(1,29) = 1.28$, MSe = 5,137.01, $p = .27$, nor the interaction, $F_1(1,29) = 0.04$, MSe = 3,612.70, $p = .84$, were significant. No interactions were significant by-items.

Errors. The data were analysed by-subjects with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subject factors, and by-items with face type, AoA and face gender as between-subjects factors. The main effect of face type was significant by-subjects, $F_1(2,58) = 6.85$, MSe = 0.70, $p<.01$, with accuracy decreasing from wholes (mean n = 0.8, 5.5%) to internals (mean n = 0.9, 6.3%) to externals (mean n = 1.2, 8.1%). Tukey tests (HSD = 0.5) indicated that there were no significant differences between any of the face types. The main effect of face type was not significant by-items, $F_2(2,168) = 1.53$, MSe = 6.23, $p = .22$. AoA was not significant in either analysis, $F_1(1,29) = 0.10$, MSe = 0.68, $p = .75$; $F_2(1,168) = 0.02$, MSe = 6.23, $p = .22$. The main effect of gender was significant by-subjects, $F_1(1,29) = 4.36$, MSe = 1.53, $p<.05$, with male faces (mean n = 0.9, 5.7%) attracting fewer errors than female faces (mean n = 1.1, 7.5%). By-items, gender was not significant, $F_2(1,168) = 2.14$, MSe = 6.23, $p = .15$.

By-subjects, two-way interactions between AoA and face type, $F_1(2,58) = 8.67$, MSe = 0.72, $p = .001$, and face type and gender, $F_1(1,58) = 5.62$, MSe = 0.52, $p<.01$, were significant. These were incorporated in a three-way interaction between AoA, gender and face type, $F_1(2,58) = 10.85$, MSe = 0.72, $p<.001$, see Figure 6.12. Data from each face type were independently scrutinized in 2 x 2 within-subject ANOVAs to probe the effect of AoA on gender.

Figure 6.12. *Relationship Between Gender and AoA as a Function of Face Type, Indexed by Gender Decision Accuracy (Experiment 8).*



In the whole condition, AoA, $F_1(1,29) = 1.45$, MSe = 0.58, $p = .24$, and gender, $F_1(1,29) = 0.21$, MSe = 0.63, $p = .65$, were not significant. However, the interaction was significant, $F_1(1,29) = 6.13$, MSe = 0.78, $p<.05$. Tukey tests (HSD = 0.6) demonstrated that none of the differences were significant. An analysis of the internals yielded significant effects of AoA, $F_1(1,29) = 5.89$, MSe = 0.46, $p<.05$, and gender, $F_1(1,29) = 3.37$, MSe = 0.99, $p = .08$, and a significant interaction, $F_1(1,29) = 6.27$, MSe = 0.77, $p<.05$. Tukey tests (HSD = 0.6) revealed a significant effect of gender on early acquired faces, with males classified more accurately than females, and a significant reversed AoA effect on female faces. The external features analysis, yielded significant effects of AoA, $F_1(1,29) = 8.31$, MSe = 1.09, $p<.01$, gender, $F_1(1,29) = 9.51$, MSe = 0.95, $p<.01$, and a significant interaction, $F_1(1,29) = 9.51$, MSe = 0.95, $p<.01$. Tukey tests (HSD = 0.7) demonstrated that gender effects were restricted to late acquired faces, with males classified more accurately than females, and the AoA effect was restricted to female external feature classification.

6.2.3.1 Participant Gender Analysis

A post-hoc analysis of the effect of participant gender was carried out as an attempt to shed light on the discrepant AoA results associated with face gender. Reaction time and error data were submitted to mixed-design by-subjects ANOVAs with face type (whole, internal or external), AoA (early or late) and face gender (male or female) as within-subjects factors and participant gender (male or female) as a between-subjects factor.

Reaction Times. Participant gender failed to modulate AoA, face type or gender.

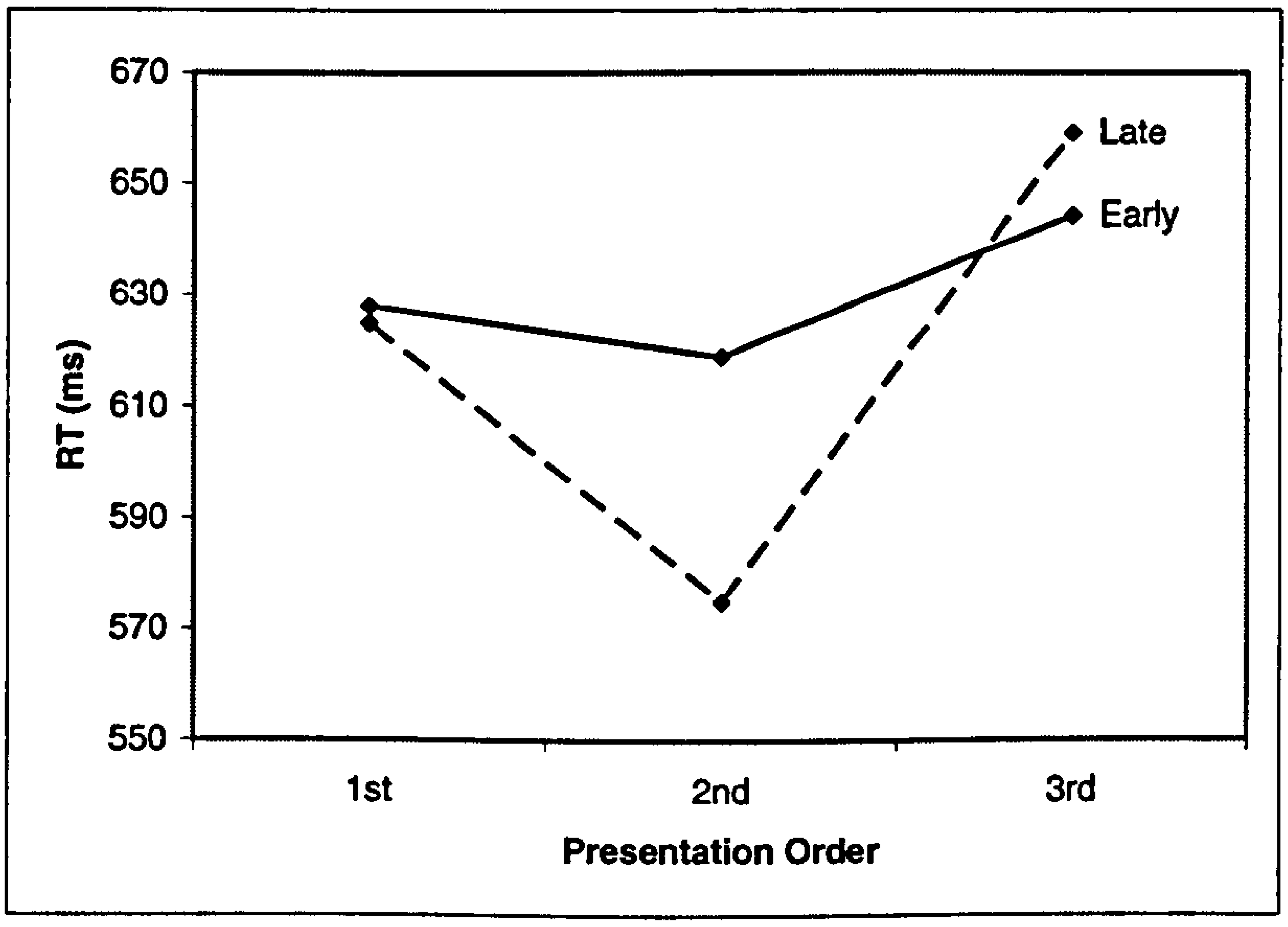
Errors. Participant gender failed to modulate face type and AoA, but interacted with face gender. In short, there was no effect of participant gender on male face classification accuracy, however female participants classified female faces significantly more accurately than male participants. Further, female participants were not affected by face gender, whilst male participants classified male faces significantly more accurately than female faces.

6.2.3.2 Presentation Order Analysis

A collection of subsidiary ANOVAs explored the effect of presentation order on each face type separately to check whether or not presentation order exerted any systematic effects. Reaction time and error data were submitted to by-subjects analyses with AoA (early or late) and gender (male or female) as within-subject factors and presentation order (first, second or third) as a between-subjects factor, and by-items analyses with AoA and gender as between-subjects factors and presentation order as a within-subjects factor.

Reaction Times. Presentation order failed to modulate effects of AoA or gender in the whole and internal feature conditions. However, the by-items analysis of externals recovered an interaction between AoA and order, $F_2(2,112) = 3.10$, MSe = 4,405.43, $p<.05$, see Figure 6.13. A simple main effects analysis demonstrated no significant effects of AoA on first, $F_2(1,58) = 0.02$, MSe = 7,844.30, $p = .90$, or third presentations, $F_2(1,58) = 0.30$, MSe = 2,421.48, $p = .59$, but AoA approached significance for second presentation data, $F_2(1,58) = 3.90$, MSe = 7,563.80, $p = .053$, although the trend was a reversal of the customary AoA effect. Further, early acquired faces were not affected by presentation order, $F_2(2,116) = 1.13$, MSe = 4,356.38, $p = .33$, but presentation order affected late acquired faces, $F_2(2,116) = 12.34$, MSe = 4,356.38, $p<.001$. Tukey tests (HSD = 40) demonstrated that second presentation data were classified significantly faster than first and third presentation data.

Figure 6.13. *Effect of AoA on Gender Decision RTs as a Function of Presentation Order (Experiment 8).*



Errors. Presentation order failed to modulate effects of gender in the internal feature condition. However, the by-subjects and by-items analyses of the whole faces yielded an interaction between gender and order. In short, female faces were classified more accurately than male faces during the first presentation, this effect reversed in the second block and by the third presentation, no gender effects emerged. Further, there was no effect of presentation order on female faces, but male faces were affected by presentation order with accuracy increasing significantly from first to second presentations, but no further differences reached significance. The by-subjects and by-items analyses of the external feature condition also yielded an interaction between gender and order: male faces were classified more accurately than female faces during the first and second presentations and this gender effect disappeared during the third presentation. Further, there was no effect of presentation order on female faces or male faces. Presentation order failed to modulate AoA regardless of face type.

6.2.3.3 Analysis of Second Presentation Groups

The variable 'presentation order' was probed further to investigate the impact of being preceded by one face type compared with another. Block 2 data were subject to scrutiny. Reaction time and error data for each face type were independently submitted to mixed-design ANOVAs. In the whole face condition, for example, data were analysed by-subjects with AoA (early or late) and gender (male or female) as within-subject factors and group (IWE or EWI) as a between-subjects factor.

Reaction Times. Group failed to modulate effects of AoA or gender, regardless of face type.

Errors. Group failed to modulate effects of AoA or gender, regardless of face type.

6.2.3.4 Familiarity Analysis

Famous face and unfamiliar face data were analysed together to provide an additional tool with which to assess the notion that gender classifications are carried out independently of identity information.

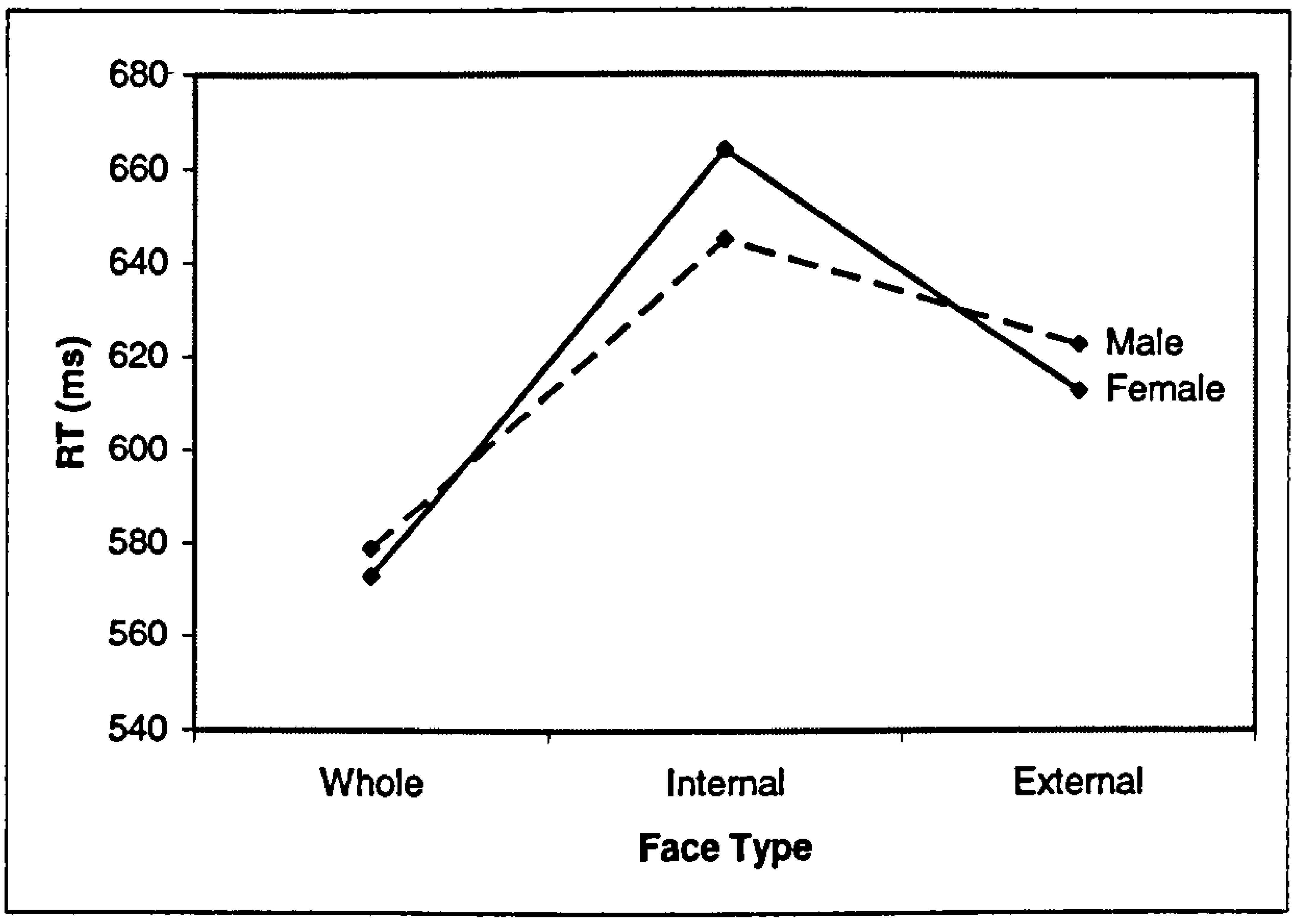
For the unfamiliar faces, the error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications made to faces correctly identified as 'unfamiliar' in the familiarity verification session; b) a failure to correctly identify a face as 'unfamiliar' in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect gender classifications, 48 (2.7%) whole face responses, 75 (4.2%) internal feature responses, and 67 (3.7%) external feature responses, were removed. In addition, 78 (4.3%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. Further, 43 (2.4%) responses from the whole condition were deemed outliers, as were 29 (1.6%) internal condition responses and 40 (2.2%) external.

Only correct gender classifications, and those made to faces correctly classified as 'unfamiliar' in the familiarity verification session, were submitted to the latency analysis. See Table 6.3 for mean unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous and unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(2,58) = 52.44$, MSe = 3,495.85, $p<.001$; $F_2(2,348) = 38.05$, MSe = 4,938.73, $p<.001$, with increasing classification latencies to wholes (576 ms), externals (618 ms) and internals (654 ms). Tukey tests ($\alpha = .05$) revealed that whole faces were classified significantly faster than external and internal features, whilst the difference between the latter two conditions reached borderline significance (HSD = 37). The main effect of familiarity was not significant by-subjects, $F_1(1,29) = 0.23$, MSe = 56,792.32, $p = .64$, but approached significance by-items, $F_2(1,348) = 3.36$, MSe = 4,938.73, $p = .068$, with faster gender classifications to unfamiliar faces (611 ms) compared with famous faces (625 ms). The main effect of gender was not significant in either analysis, $F_1(1,29) = 0.02$, MSe = 2,510.89, $p = .89$; $F_2(1,348) = 0.09$, MSe = 4,938.73, $p = .77$.

A two-way interaction between face type and gender was significant in the by-subjects analysis, $F_1(2,58) = 4.50$, MSe = 1,717.25, $p<.05$, see Figure 6.14. A simple main effects analysis demonstrated that the classification of whole faces, $F_1(1,29) = 0.62$, MSe = 876.05, $p = .44$, and external features, $F_1(1,29) = 1.44$, MSe = 1,235.84, $p = .24$, was not affected by face gender. Internal feature classifications were modulated by participant gender, $F_1(1,29) = 6.24$, MSe = 865.40, $p<.05$, with males classified faster than females. Moreover, the simple main effects analysis revealed that responses to female faces, $F_1(2,58) = 45.07$, MSe = 1,372.20, $p<.001$, and male faces, $F_1(2,58) = 27.25$, MSe = 1,232.91, $p<.001$, were modulated by face type. Tukey tests demonstrated that responses to each face type condition were significantly different. This was true for female faces (HSD = 23) and male faces (HSD = 22). No significant interactions emerged from the by-item analysis.

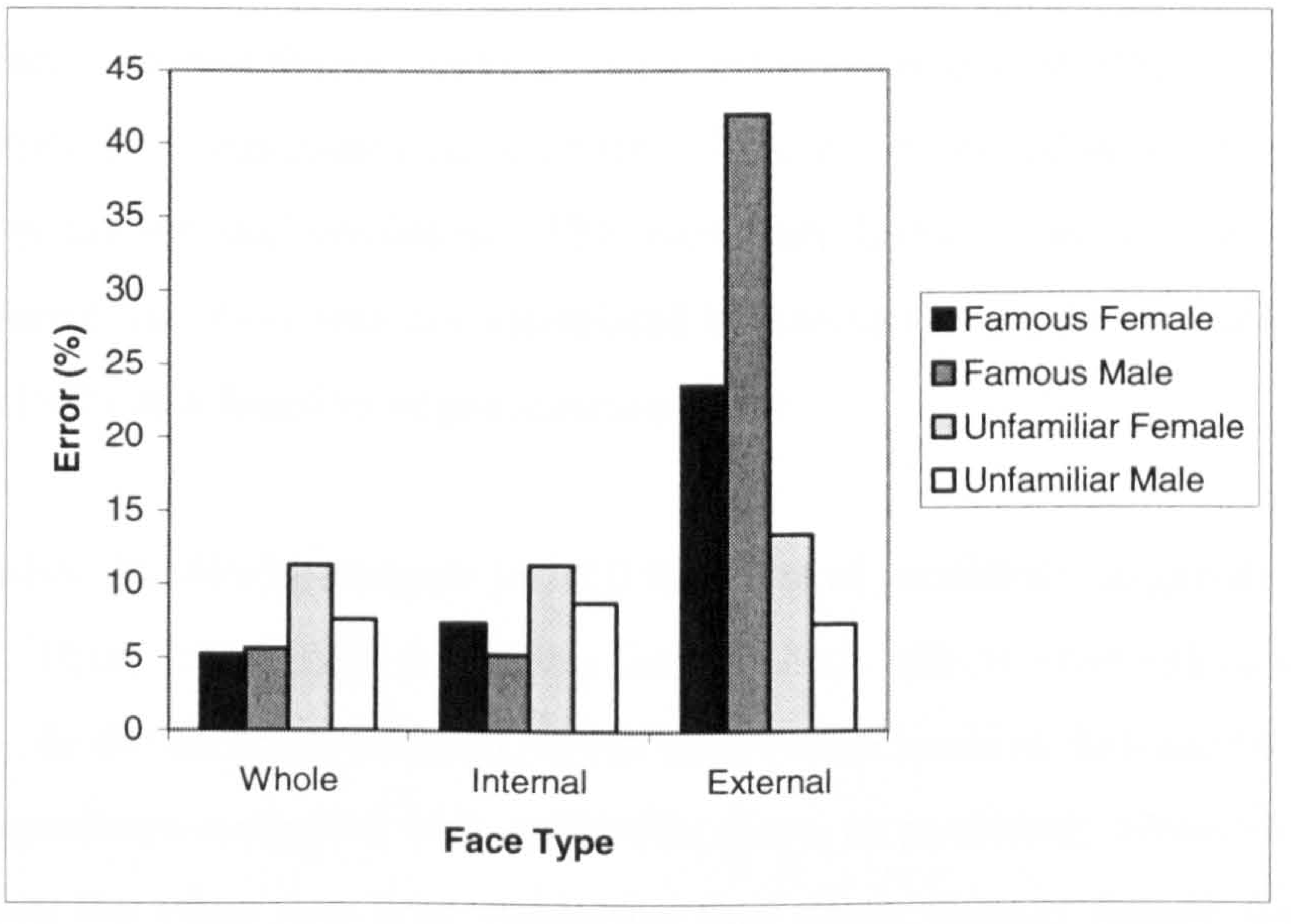
Figure 6.14. *Effect of Face Type on Gender Decision RTs as a Function of Gender (Experiment 8).*



Errors. The data were analysed by-subjects with face type (whole, internal or external), familiarity (famous and unfamiliar) and face gender (male or female) as within-subjects factors, and by-items with face type, familiarity and face gender as between-subjects factors. The main effect of face type was significant by-subjects, $F_1(2,58) = 70.28$, MSe = 9.67, $p<.001$, with increasing errors to wholes (mean n = 2.2, 7.5%), internals (mean n = 2.5, 8.2%) and externals (mean n = 6.5, 21.6%). Tukey tests (HSD = 1.9) indicated that externals were classified significantly less accurately than wholes and internals, but the latter conditions were not significantly different. By-items, face type was not significant, $F_2(2,348) = 1.38$, MSe = 6.02, $p = .25$. The main effect of familiarity was significant in both analyses, $F_1(1,29) = 6.53$, MSe = 29.20, $p<.05$; $F_2(1,348) = 15.62$, MSe = 6.02, $p<.001$, with famous faces (mean n = 4.5, 14.8%) attracting more errors than unfamiliar faces (mean n = 3.0, 10%). The main effect of gender was not significant by-subjects, $F_1(1,29) = 0.93$, MSe = 4.32, $p = .34$, but was significant by-items, $F_2(1,348) = 11.52$, MSe = 6.02, $p = .001$, with fewer errors made to male faces (mean n = 2.1, 6.9%) compared with female faces (mean n = 2.9, 9.8%).

Significant two-way interactions between face type and familiarity, $F_1(2,58) = 80.25$, MSe = 7.71, $p<.001$, face type and gender, $F_1(2,58) = 39.78$, MSe = 1.50, $p<.001$, and familiarity and gender, $F_1(1,29) = 54.34$, MSe = 3.46, $p<.001$, emerged from the by-subjects analysis. These were subsumed in a significant interaction between face type, familiarity and gender, $F_1(2,58) = 43.76$, MSe = 2.59, $p<.001$, see Figure 6.15. Data from each face type condition were independently scrutinized in 2 x 2 within-subject ANOVAs to probe the effect of familiarity on gender.

Figure 6.15. *Relationship Between Gender and Familiarity as a Function of Face Type, Indexed by Gender Decision Accuracy (Experiment 8).*



In the whole condition, familiarity was significant, $F_1(1,29) = 8.71$, MSe = 5.10, $p<.01$, and gender approached significance, $F_1(1,29) = 3.89$, MSe = 1.80, $p = .058$. The interaction was significant, $F_1(1,29) = 7.55$, MSe = 1.51, $p = .01$. Tukey tests (HSD = 0.9) revealed that gender affected unfamiliar faces with male faces classified more accurately than female faces, and familiarity effects were restricted to female faces with fewer errors made to famous female faces than unfamiliar female faces. In the internal condition, familiarity, $F_1(1,29) = 5.76$, MSe = 6.50, $p<.05$, and gender, $F_1(1,29) = 6.64$, MSe = 2.32, $p<.05$, were significant. The interaction was not significant, $F_1(1,29) = 0.02$, MSe = 3.47, $p = .88$. In the external condition, familiarity, $F_1(1,29) = 40.78$, MSe = 33.03, $p<.001$, and gender, $F_1(1,29) = 31.56$, MSe = 3.20, $p<.001$, were significant. Further, the interaction was significant, $F_1(1,29) = 110.17$, MSe = 3.66, $p<.001$. Tukey tests (HSD = 1.4) demonstrated that all differences were significant: for both male and female faces, unfamiliar faces were classified more accurately than famous faces; famous female faces were classified more accurately than famous male faces, and unfamiliar faces showed the reversed gender effect.

6.2.4 Discussion

The latency analysis yielded a reversed AoA effect on male faces in the whole face condition and the accuracy analysis yielded a disappointing spread of inconsistent AoA effects: internal female features exhibited a reversed AoA effect and external female features, a customary AoA effect. Thus, the predicted AoA effect did not emerge in the internal condition. The subsidiary latency and accuracy analyses demonstrated that AoA was not modulated by participant gender and did not vary systematically as a function of presentation order.

The post-hoc familiarity analysis yielded no effect of familiarity on gender decision latency. Thus, it is hardly surprising that no AoA effects emerged in the main analysis. In the accuracy analysis, fewer errors were made to famous faces in the internal condition compared with unfamiliar faces, as predicted. However, effects surfaced in the other face type conditions too: whole famous female faces were classified more accurately than whole unfamiliar female faces and in the external condition, a reversed familiarity effect emerged.

Both the Experiment and the post-hoc familiarity analysis demonstrated inconsistent AoA effects. Given that the familiarity decision task (Experiment 7) failed to yield reliable AoA effects across all face types, these disappointing effects were not surprising. Whilst Experiments 7 and 8 failed to replicate Experiments 5 and 6, they served to demonstrate that discrepant AoA effects associated with face gender were likely to be artefacts of the stimulus set and not an intrinsic characteristic of face processing, necessitating further exploration.

6.3 General Discussion: Experiments 5 to 8

Experiments 5 to 8 sought to ascertain the nature of the relationship between identity and gender processing and determine the conditions under which identity affects gender processing. In Experiment 6, AoA effects in the male internal feature condition were taken to be an index of this relationship between the cognitive operations. However, this finding was somewhat overshadowed by inconsistent results from one experiment to the next and discrepancies between male and female faces within experiments. Therefore, conclusions are not clear-cut. For the sake of brevity and clarity, accuracy analyses are not subject to discussion; suffice to say that speed-accuracy trade-offs were not operating to compromise AoA effects in the latency analyses. Moreover, recall the limited validity of accuracy analyses: misclassifications of target faces unfamiliar to participants cannot be discriminated from misclassifications of known faces[21].

Experiments 5 and 6 provided some evidence for an effect of identity on gender processing. In the familiarity decision task, performance conformed to expectation with AoA affecting whole face, internal feature and external feature recognition. However, these effects were restricted to male faces. In light of this, AoA was exclusively predicted to modulate gender decisions made to male faces. This prediction was confirmed leading to the conclusion that this relationship between the cognitive operations arises when gender decisions cannot be made on the basis of superficial gender cues alone. Unfortunately, these results were not replicated. In the second familiarity decision task, recognition of internal features was not affected by AoA. Moreover, effects of AoA on whole face and external features were restricted to female faces. These inconsistent effects mapped onto the gender classification task with AoA exclusively affecting whole male faces. It is logical to suppose that effects of AoA on gender processing of internal features must be preceded by robust effects in a familiarity decision task.

---

21 Note that the summary results tables for each experiment report '% misclassification' which denotes errors made to faces which had been correctly recognized in the familiarity verification session, in addition to '% errors'. A visual inspection of both measures suggests that they both display the same trends. Thus, the accuracy analysis carried out on '% errors' may be of limited validity, but the results would probably not have been too different if carried out on '% misclassifications'.

Whilst the replicated experiments failed to support the pivotal finding, they served to dispel the possibility that AoA exclusively affects male face recognition performance. Rather, the discrepant gender effects in Experiments 5 and 6 appeared to be an artefact of the stimulus set. Unfortunately, the erratic nature of the AoA effect throughout Experiments 5 to 8 suggests that the stimulus sets employed may not be optimally suited to the task of recovering AoA effects on gender processing.

Post-hoc familiarity analyses comparing gender decisions made to famous and unfamiliar faces, provided an additional assessment of the relationship between identity and gender processing. Familiarity decisions made to famous and unfamiliar faces (Experiments 5 and 7) did not vary as a function of familiarity, contrary to predictions and the findings of Bruce et al. (1987a). Given the problems inherent in comparing responses to stimulus categories which correspond to separate binary decisions (Brysbaert et al., 2000), this discrepancy was not disconcerting. In the first gender decision task (Experiment 6), judgments of famous internal features were facilitated relative to unfamiliar internals. Whilst this matched predictions, interpretation of this finding was somewhat undermined by the finding that external feature classification also varied as a function of familiarity and that all familiarity effects stemmed from female faces. However, these analyses should be treated with caution because the famous and unfamiliar sets were not matched on pertinent factors such as gender typicality and distinctiveness. Further, no familiarity effects stemmed from the post-hoc analysis of Experiment 8.

Both parallel-route and single-route hypotheses offer explanations for the effect of identity on gender processing found in Experiment 6. The semantic access hypothesis is one explanation favoured by advocates of parallel-route models. However, Bruce (1986) suggested that the time-courses of both operations must be comparable to allow semantic information to facilitate gender decisions. Given that the gender decision latencies in Experiment 6 were sufficiently quick relative to the corresponding familiarity decisions in Experiment 5, this renders the semantic access hypotheses an unlikely candidate. Moreover, the only item with overlapping response times was a set of internal features belonging to Michael Jackson! Clutterbuck and Johnston's (2004) hypothesis bypasses problems associated with timely semantic access. Their suggestion that familiar faces use FRU feedback as well as structurally-encoded information to enable gender processing does provide one fitting explanation of the current data.

McNeil et al. (2003) collected persuasive evidence in favour of a semantic access hypothesis. They demonstrated that when participants are coerced to use semantic information in order to make gender judgments, for example deciding the gender of celebrities on the basis of surnames, an effect of identity on gender decisions emerges. Nonetheless, parallel-route accounts do require modification to alleviate problems associated with the relative time-courses of the operations. Rossion's (2002) cascade mechanism is one such modification which allows semantics to aid gender decisions even when the latter is processed more quickly than the former. Again, this mechanism could account for the affect of identity on gender processing in Experiment 6.

Goshen-Gottstein and Ganel's (2000) single-route hypothesis provides a deceptively parsimonious explanation of the current data. However, the single-route hypothesis relies on acceptance of three rather contentious sub-hypotheses. First, FRUs sub-serve gender classification tasks and a single exposure of a face is sufficient for their formation. Second, repetition priming results from the reactivation of representations stored in FRUs, contrary to proponents of recent parallel-route accounts. Third, integrality of operations will not arise when participants 'truncate the perceptual whole' by adopting superficial hair-style heuristics to make gender decisions. The authors specifically predicted that an integrality of identity and gender processing will only emerge when participants base gender judgments on hair-deleted stimuli. This third sub-hypothesis can be exploited to establish whether parallel-route or single-route accounts are the most apt to explain the present result.

A post-hoc investigation of the gender classification data from Experiment 6 attempted to identify the most likely explanation of the results. The size of AoA effects on gender decisions made to faces which had previously displayed relatively fast familiarity decisions was compared with the size of effects stemming from faces with relatively slow familiarity decisions. As the single-route hypothesis only predicts an integrality of processing when hair-deleted stimuli are used, according to this hypothesis it is unlikely that insertion of this factor would modulate the results of Experiment 6. Conversely, parallel-route accounts predict an effect of identity on gender processing whenever the processing of sexually dimorphic features is laborious; whether faces are androgynous or if external features are removed. Thus, such accounts would predict that faces with relatively slow gender decisions and fast familiarity decisions are more likely to exhibit the effect than faces with the converse pattern of responses. Whilst gender decisions to internal features were not modulated by the relative time-courses of the operations, responses to whole faces were affected: faces with relatively slow gender decisions and fast familiarity decisions showed a tendency for early acquired wholes to be processed faster than late acquired wholes. Although these results are not particularly compelling, the latter result is compatible with mechanisms associated with parallel-route accounts. To further probe this investigation, Experiments 9 and 10 address the impact of gender typicality on the relationship between identity and gender processing.

Evidence in favour of a parallel-route account also stems from the contrasting nature of the perceptual representations extracted during identity and gender processing. Whilst the 3D structure or facial protuberance associated with certain cues is important for gender processing (Bruce et al., 1993a; Burton et al., 1993; Bruce & Langton, 1994; Enlow, 1982), this information source appears to be less salient for identification (Bruce & Langton, 1994; Bruce et al., 1987b). Had the perceptual representations for identity and gender processing been identical, then it is conceivable how familiarity with a face could facilitate the extraction of information important for identity and gender decisions. Moreover, a single processing route would have provided a parsimonious explanation of this relationship. As it stands, the contrasting representations or cues associated with each process are more in keeping with an account which advocates functional independence.

The inconsistent findings in these experiments makes it difficult to pinpoint particular weaknesses in the experimental design, suffice to say that artefacts associated with the stimuli may have provoked disappointing results. As the databases had been created approximately two years prior to these experiments, the stimuli may not have been optimal for use with the particular cohort of students tested: the transient nature of faces exacerbates the task of finding a large enough sample of sufficiently familiar stimuli.

In summary, where robust AoA effects modulated familiarity decisions to internal features (Experiment 5), these translated into effects on gender decisions (Experiment 6). From this result, it appears that the mechanisms underpinning identity and gender processing are not entirely independent. Additionally, the gender decision task provided a means of measuring AoA effects in a task which did not demand explicit recognition of target faces, and thus AoA effects were less vulnerable to distortion through participant error. Experiments 7 and 8 failed to replicate this effect but this did not constitute compelling evidence to dissuade the hypothesis that identity affects gender processing. It is anticipated that Experiments 9 and 10 will not only extend the present findings, but consolidate existing uncertainties.

CHAPTER 7 – FURTHER EXPLORATIONS INTO THE RELATIONSHIP BETWEEN IDENTITY AND GENDER PROCESSING: RECRUITING AOA AS AN INVESTIGATIVE TOOL

Experiments 9 and 10 constituted a further attempt to ascertain the nature of the relationship between identity and gender processing. Gender typicality (masculinity / femininity) was inserted into the design to distinguish single-route and parallel-route hypotheses. Whereas parallel-route accounts predict an effect of identity on gender processing whenever processing via the visually-derived semantic route is slowed, the single-route account hypothesises that integrality emerges specifically when external gender cues are removed; when the perceptual whole is truncated. An investigation of the effect of gender typicality on AoA should serve to delineate mechanisms underpinning the relationship between identity and gender processing.

Gender typicality ratings were collected and a post-hoc analysis of Experiment 6 was carried out with gender typicality inserted retrospectively as a factor. This analysis explored the possibility that gender typicality could modulate the relationship between identity and gender processing before committing to Experiments 9 and 10.

7.1 Collection of Gender Typicality Ratings

Famous and unfamiliar whole faces and internal features were rated for gender typicality with the famous images taken from the current database. The 253 whole faces were rated by 22 York University students (mean age = 19.77 years old, S.D. = 2.39). The corresponding internal features were rated by 21 participants from the same population (mean age = 18.57 years old, S.D. = 0.75). The tasks were administered as booklets with a cover page detailing instructions and a rating scale printed on the top of each page. Participants were asked to complete the task in the order specified. The 253 images were spread over 12 pages and items were pseudo-randomised between participants in that the pages were ordered differently. Famous and unfamiliar faces were intermixed, but male and female faces were blocked and the order of the blocks was counterbalanced. Participants were requested to rate the faces according to femininity or masculinity, using a 7-point scale (1 = not very masculine / not very feminine, 7 = highly feminine / highly masculine).

The split-half reliability of whole face and internal feature ratings was checked by correlating the mean ratings of half of the participants with the mean ratings of the remaining half. For both face types, $r = .90$, indicating highly reliable ratings.

Correlations between gender typicality for the famous whole faces and internal features, and previously collected ratings of AoA, distinctiveness, frequency, likeness and familiarity are shown in Table 7.1. The analyses were performed using two-tailed Pearson's Correlation Coefficients. Eight significant correlations emerged from the analysis but six of these concerned previously rated variables (see p. 77) and so will not be reported again. A strong positive correlation between whole face and internal feature gender typicality emerged, $r(122) = .82$, $p<.001$. Moreover, a paired t-test indicated that neither face type received significantly different ratings, $t(123) = -.10$, $p = .92$. This suggests that the presence of external features, in the whole face condition, does not change the perceived degree of gender typicality, relative to internal feature ratings, despite speeding up gender classifications (see Experiments 6 and 8).

Table 7.1. *Correlation Matrix for the 7 Variables in the Analysis of Current Images (n = 124).*

	AoA	Distinctive	Frequency	Likeness	Familiarity	Whole: Gender Typ.	Internal: Gender Typ.
AoA	1.00						
Distinctive	-0.16	1.00					
Frequency	0.51**	-0.07	1.00				
Likeness	0.04	0.18*	0.23*	1.00			
Familiarity	0.13	0.24**	0.46**	0.32**	1.00		
Whole: Gender Typ.	0.18	-0.04	0.12	0.11	0.19*	1.00	
Internal: Gender Typ.	0.10	0.12	-0.03	0.06	0.12	0.82**	1.00

*. Correlation is significant at the 0.01 level (two-tailed)
*. Correlation is significant at the 0.05 level (two-tailed)

A very weak correlation emerged between whole face gender typicality and familiarity, indicating that faces rated as familiar tended to be rated as gender typical, $r(122) = .19$, $p<.05$. Although it is hard to conceive how familiarity can impact on structural facial characteristics, it is possible that having a gender typical face makes it easier to become famous. Perhaps this social factor is responsible for this correlation.

7.2 Post-Hoc Analysis of Experiment 6 with Gender Typicality as a Factor

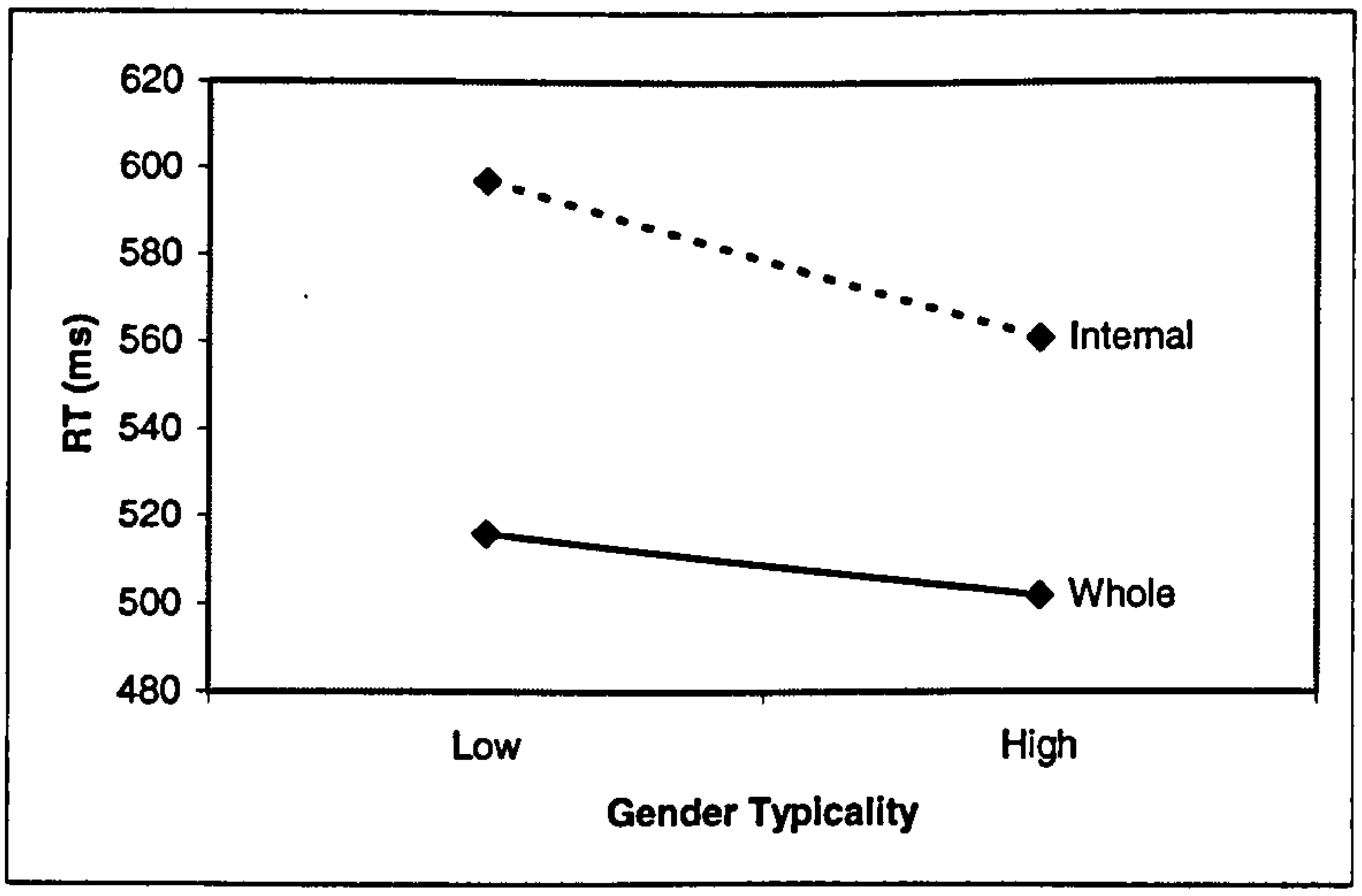
Experiment 6 yielded an effect of AoA on gender classifications made to internal features. However, this effect was restricted to male faces with a reverse AoA effect for female internals. A post-hoc exploration of the latency analysis assessed whether or not gender typicality modulated this interaction. Whole face and internal feature items were each split into high and low gender typicality sets, determined by median gender typicality. Table 7.2 reports the mean correct response latencies with gender typicality inserted as a factor.

Table 7.2. *Mean RTs for Gender Decisions in Each Condition (Post-Hoc Analysis of Experiment 6).*

			Early Acquired			Late Acquired		
Stimulus			Male	Female	Mean	Male	Female	Mean
Whole	Low Gender Typ.	*Mean RT*	522	500	510	532	509	524
		SD	82	80	75	75	100	75
	High Gender Typ.	*Mean RT*	511	496	504	501	500	500
		SD	88	89	84	77	75	70
Internal	Low Gender Typ.	*Mean RT*	592	599	596	620	575	601
		SD	103	98	84	95	105	89
	High Gender Typ.	*Mean RT*	570	551	563	577	548	559
		SD	73	97	78	97	81	81

The data were analysed by-subjects with face type (whole or internal), AoA (early or late), face gender (male or female) and gender typicality (high or low) as within-subjects factors. For the sake of brevity, only those findings which implicate gender typicality are reported (the other results do not differ from those reported in the original analysis of Experiment 6). The main effect of gender typicality was significant, $F_1(1,35) = 20.78$, MSe = 4,183.97, $p<.001$, with gender typical faces (532 ms) classified according to gender faster than faces low in gender typicality (556 ms). Also, gender typicality interacted significantly with face type, $F_1(1,35) = 5.84$, MSe = 2,805.61, $p<.05$, see Figure 7.1. Tukey tests (HSD = 34) demonstrated that whole faces were classified faster than internal features regardless of gender typicality. Further, gender typicality affected internal feature classification in the customary direction but whole face performance was not modulated. The interaction indicates that external features mask effects of gender typicality implying, somewhat speculatively, that these features do supply superficial gender information to facilitate gender decisions.

Figure 7.1. *Effect of Gender Typicality on Gender Classification RTs as a Function of Face Type (Post-Hoc Analysis of Experiment 6).*



Although gender typicality did not interact significantly with AoA, visual inspection of Table 7.2 shows several trends. For example, male internals low in gender typicality yielded an AoA effect of 28 ms, compared with a marginal effect of 7 ms for gender typicality internals. Thus, there appears to be some evidence to warrant further investigation of the effect of gender typicality on the relationship between identity and gender processing.

7.3 Introduction: Experiments 9 and 10

At present, a common trait of the face recognition literature is the erratic way in which the various parallel-route and single-route accounts are employed to explain the relationship between identity and gender processing: parallel-route and single-route mechanisms appear to be effectively interchangeable across studies. Experiments 9 and 10 sought to identify possible mechanisms underpinning the effect of identity on gender processing found in Experiment 6 to impose some theoretical constraints on research in this area. The experimental manipulations were designed to exploit discrepancies between predictions made by the parallel-route and single-route accounts, and by so doing, to identify the most capable candidate for explaining the relationship between identity and gender processing. Studies typically remove external cues to gender, such as hair-style and face-shape, to invoke stronger effects of identity on gender processing. Moreover, all studies argue that the presence of external gender cues reduces the opportunity for identity information to influence the decision. However, the parallel-route and single-route accounts make different predictions regarding the impact of gender typicality. Experiments 9 and 10 manipulated gender typicality in order to disentangle the various explanations. In order to pit the two accounts against one another, the predictions made by each one are first extrapolated to derive predictions regarding the impact of gender typicality.

Goshen-Gottstein and Ganel (2000) and Ganel and Goshen-Gottstein (2002) maintained that integrality between cognitive operations is the default in face recognition systems, and false separability only emerges when participants base gender decisions on hair-style heuristics. The mechanism for this shift in processing is described by the truncated processing hypothesis. This deems that adequate processing of the perceptual whole is vital for engaging FRUs, which mediate both identity and gender processing. The authors argued that critical processing of internal features is overridden specifically by hair-style heuristics and as a result, integrality of processing only emerges when external features are removed. Conversely, parallel-route accounts argue that the cognitive operations are functionally and neuroanatomically distinct but appear interactive *whenever* processing via the visually-derived semantic route is slowed. Clutterbuck and Johnston (2004) and Rossion (2002) induced interactivity by removing external cues to gender whereas Bruce (1986) speculated that interactivity arose as a result of processing androgynous faces.

Most studies investigating the relationship between identity and gender processing vary the ease of gender processing by removing external cues. Inevitably, effects can be interpreted as support for single-route or parallel-route hypotheses. Intrinsic physiognomic variations in gender typicality are seldom brought to bear on this subject, with the exception of Bruce's (1986) study. This is particularly surprising since the removal of external features has limited external validity compared with variations in gender typicality. By assessing the impact of gender typicality on the relationship between the cognitive operations, it may be possible to disentangle single-route and parallel-route accounts. Given that the single-route account is driven by the truncated processing hypothesis, which is not sympathetic to variations in gender typicality, it is logically inferred that the single-route account predicts that effects of identity on gender processing should not be modulated by gender typicality. It is also inferred that parallel-route accounts predict the relationship between identity and gender to vary as a function of gender typicality: mechanisms underpinning these accounts are sympathetic to any experimental manipulations which make gender processing difficult.

There are several reasons why the single-route account is, *a priori*, not favourable. It appears parsimonious upon first inspection but encounters several pragmatic and theoretical limitations, particularly when recruited to explain repetition priming onto a gender decision (Ganel & Goshen-Gottstein, 2000). Firstly, Ganel and Goshen-Gottstein couch the hypothesis in terms of memory systems accounts (e.g. Moscovitch, Goshen-Gottstein & Vriezen, 1993) which are unable to explain the cross-domain (face/name) priming effects found in McNeill et al.'s (2003) study. Secondly, to explain priming with unfamiliar faces, Ganel and Goshen-Gottstein hypothesised that repetition priming results from reactivation of representations stored in FRUs, contrary to recent accounts, and inferred that a single presentation of a face must be sufficient to form an FRU. However, Clutterbuck and Johnston

(2004) demonstrated that even 10 exposures of an unfamiliar face failed to form an FRU comparable to that of a familiar face. The constraints imposed by the single-route hypothesis seem to be post-hoc attempts to explain the data, which do not sit comfortably with other studies.

In light of this, it was hypothesised that gender typicality would modulate effects of AoA on gender processing, providing support for parallel-route hypotheses. Gender decisions were made to full faces and internal features split factorially according to AoA and gender typicality. Experiments 6 and 8 demonstrated that external features can support gender decisions and are therefore capable of cueing gender: this did not need to be demonstrated once again. To ensure that the stimulus set employed is capable of yielding AoA effects, a familiarity decision task (Experiment 9) was administered prior to the gender decision task (Experiment 10). Both experiments were followed by a familiarity verification session for participants to indicate which faces were known to them.

AoA was hypothesised to affect familiarity decisions made to whole faces and internal features, and gender typicality was predicted not to modulate these effects. Bruce et al. (1987a) recovered no effects of gender typicality on familiarity decisions and as Clutterbuck and Johnston (2004) put it, "it is difficult to predict any circumstances where information from the gender process could ever assist determining the identity of a face" (p. 166). In the gender decision task, faces low in gender typicality were predicted to elicit AoA effects regardless of face type. Conversely, gender typical faces were predicted to show AoA effects in the internal condition only, and it was anticipated that these effects would be slightly subdued compared with those elicited by low gender typicality internals.

A post-hoc familiarity analysis was carried out to offer further evidence that familiarity can influence gender processing and that this is modulated by gender typicality. It was hypothesised that gender decisions to famous faces would be facilitated relative to unfamiliar faces for faces low in gender typicality, regardless of face type. Effects of familiarity on gender typical faces were predicted to emerge in the internal feature condition only.

7.4 Experiment 9

7.4.1 Introduction

Experiment 9 investigated the effect of AoA on familiarity decisions made to whole faces and internal features, further subdivided into high and low gender typicality items. This ensured that the stimulus set readily yielded AoA effects as a prerequisite to Experiment 10. Stimuli were taken from the current database because the experiments using dated images (Experiments 7 and 8) failed to yield promising insights into the relationship between identity and gender processing. Owing to limitations in the size of the database, experimental stimuli comprised male faces only with female faces serving as filler items. A familiarity verification session administered after the experiment proper ensured that the latency analysis was based only on responses made to faces genuinely recognised by participants.

It was hypothesised that AoA would affect familiarity decisions regardless of face type and gender typicality. In light of the results from the post-hoc familiarity analyses of Experiments 5 and 7, where famous and unfamiliar faces had similar classification latencies, it was predicted that familiarity would not affect familiarity decisions: this analysis was carried out to facilitate comparisons with the post-hoc familiarity analysis of Experiment 10. However, its value is limited because it is suboptimal to compare responses to categories which correspond to separate responses in a binary decision task (Brysbaert et al., 2000). Also, the accuracy analysis is limited because it taps a bias for participants to respond 'unfamiliar' when uncertain, resulting in a greater number of errors to famous faces compared with unfamiliar faces.

7.4.2 Method

Participants. Thirty two undergraduate students from the University of York took part in this experiment (5 male, 27 female), receiving course credit or a payment of £2 for their participation. Participants were 19 – 22 years old (mean = 20.13; S.D. = 0.91) and were required to have lived in the UK for at least 18 years.

Design. The experimental variables comprised face type (with two levels; whole face or internal features), AoA (with two levels; early or late) and gender typicality (with two levels; high or low) which were explored in a repeated measures design. The order of the face types was counterbalanced across participants.

The dependent variables, response latency and accuracy, were the products of a familiarity decision task.

Materials. Forty male faces (10 early acquired gender typical, 10 early acquired low gender typicality, 10 late acquired gender typical, 10 late acquired low gender typicality) were selected from the current database. Independent t-tests confirmed that the four sets were matched for frequency, distinctiveness and likeness, that the two early sets and two late sets were matched for AoA, and that the two gender typical and two low gender typicality sets were matched for gender typicality (in each case $t \leq 1.37$), see Table 7.3 for summary data for the experimental sets.

Achieving a factorial split of gender typicality in conjunction with the manipulation of other variables was difficult, given the limited size of the database. Hence the limited set sizes and the less than ideal matching of sets: the lowest gender typicality rating in the early acquired gender typical set equals the highest gender typicality rating in the late acquired low gender typicality set. This compromise was unavoidable in view of the limited resources. Further, it was not possible to match female sets in accordance with all the required characteristics, again owing to the limited size of the database. Therefore, 40 coarsely matched female faces (10 early acquired gender typical, 10 early acquired low gender typicality, 10 late acquired gender typical, 10 late acquired low gender typicality) were selected to act as filler items. Although not necessary for this task, the inclusion of filler items served to facilitate the forthcoming gender decision task.

Faces in the internal condition were edited in the same manner as those featured in Experiments 5 to 10, and examples of whole face and internal feature stimuli are shown on p 141. Experimental set composition was determined by whole face gender typicality ratings. Given the strong correlation between whole face and internal feature gender typicality ratings this was not considered detrimental to the design.

The experimental items and filler items were matched with an equal number of unfamiliar faces on gender, approximate age and racial group. Furthermore, efforts were made to match the famous and unfamiliar male faces according to gender typicality (in each case, $t \leq 1.44$), see Table 7.4 for famous and unfamiliar set summary data. By matching gender typicality across famous and unfamiliar sets, it was intended that the quality of the post-hoc familiarity analysis in Experiment 10 would be improved[22]. This measure was a feature of Experiment 9 by virtue of the fact that both experiments use the same stimulus set.

---

22 A post-hoc analysis of Experiment 6 demonstrated that famous and unfamiliar whole face sets had been matched for gender typicality, $t(134) = 1.66$, $p = .10$, as had internal sets, $t(134) = 1.25$, $p = .21$. Although this validated the finding that famous internals were classified according to gender faster than unfamiliar internals, a post-hoc insertion of gender typicality as an additional factor demonstrated that low gender typicality famous faces had significantly higher ratings than low gender typicality unfamiliar faces. This was true for whole faces, $t(66) = 3.43$, $p<.001$ and internals, $t(66) = 2.86$, $p<.01$. However, famous and unfamiliar gender typical sets did not have different ratings. Evidently if gender typicality is left unmatched, there are circumstances when it may vary between famous and unfamiliar sets. Hence, the decision to match famous and unfamiliar sets for gender typicality in Experiments 9 and 10.

Table 7.3. *Mean Ratings as a Function of Experimental Set (Experiment 9).*

	Early acq. & low gender typ.			Early acq. & high gender typ.			Late acq. & low gender typ.			Late acq. & high gender typ.		
Rating	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
AoA	2.92	(0.42)	2.14 - 3.41	2.82	(0.37)	2.14 - 3.25	4.23	(0.55)	3.62 - 5.00	3.94	(0.41)	3.57 - 4.95
Gender typ.	3.35	(0.76)	1.32 - 3.91	4.81	(0.43)	4.23 - 5.58	3.36	(0.82)	2.05 - 4.23	4.75	(0.52)	4.27 - 5.64
Frequency	2.46	(0.56)	1.80 - 3.65	2.49	(0.34)	2.05 - 3.15	2.75	(0.56)	1.85 - 3.70	2.50	(0.37)	1.90 - 3.15
Distinctiveness	2.97	(1.06)	1.65 - 4.80	2.97	(0.68)	1.90 - 3.90	2.94	(0.71)	1.80 - 4.13	2.83	(0.76)	1.60 - 3.90
Likeness	3.95	(0.46)	3.00 - 4.62	3.93	(0.22)	3.64 - 4.21	3.82	(0.36)	3.38 - 4.52	4.01	(0.35)	3.52 - 4.71

Table 7.4. *Mean Gender Typicality Ratings as a Function of Famous and Unfamiliar Sets (Experiment 9).*

	Famous gender typ.			Unfamiliar gender typ.			Famous low gender typ.			Unfamiliar low gender typ.		
Face type	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*	*Mean*	*(SD)*	*Range*
Whole	4.78	(0.47)	4.23 - 5.64	4.97	(0.38)	4.45 - 5.82	3.35	(0.77)	1.32 - 4.23	3.59	(0.75)	2.27 - 4.45
Internal	4.82	(0.64)	3.43 - 5.71	4.69	(0.75)	3.10 - 5.95	3.67	(0.97)	1.38 - 4.95	3.57	(0.80)	2.24 - 5.14

Procedure and Apparatus. The familiarity decision task was executed in PsyScope (Cohen et al., 1993) for the AppleMac. Participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is familiar" in response to images presented in the centre of the screen. Each experimental trial began with a 500 ms blank screen prior to image onset. Faces were displayed until a response was made. 'Familiar' and 'unfamiliar' responses were logged using the 'z' and '/' keys on the keyboard, and keys pressed with index fingers. The labelling of keys was counter-balanced across participants. The images were displayed on a screen with a resolution of 1,280 x 1,024 pixels and the absolute dimensions of the whole face images were approximately 70 x 150 mm. Participants viewed these from a distance of approximately 30cm.

Experimental trials and filler items were presented once as whole faces and once as internal features. The face type conditions were blocked, and the order of the blocks was counterbalanced. Each face type block was further subdivided into two blocks, and the order of these sub-blocks randomized between participants. This subdivision was introduced into the design to reduce the likelihood of associated famous faces appearing in succession, causing associative priming.

Each block was preceded by task instructions and information about the nature of the stimuli. For example, an excerpt from the internal condition instructions read "you are about to see a series of PART celebrity faces and unfamiliar faces presented one at a time. The parts will comprise internal features (eyes, nose, mouth). See example below". The instructions preceding the second block varied slightly from those read at the start as participants were additionally informed that the familiarity decision referred specifically to whether or not the face was famous. This line was inserted to counteract a growing sense of familiarity for the unfamiliar faces stemming from their repeated presentation over the two blocks.

Thirty two practice items (16 famous and 16 unfamiliar) were administered prior to each of the blocks. After the practice session, participants were reminded to respond "as quickly and accurately as possible". The experimental trials were directly preceded by 8 lead-in trials. A representative sample of stimuli (early and late acquired; high and low gender typicality; male and female faces) constituted practice and lead-in items. None of these were repeated as experimental items or included in the analysis. Practice and lead-in images were presented as whole faces or internal features as appropriate.

After the experiment, participants completed a familiarity verification exercise. Participants were presented with the intact versions of the 160 faces and asked to indicate at their own pace "which faces are familiar to you (i.e. you recognize them as being famous)". They were explicitly informed that the task was not timed and accuracy was paramount. Coupled with the experiment proper, the session lasted approximately 25 minutes.

7.4.3 Results

Eleven participants were replaced after failing to meet the criteria which required them to: a) produce 75% correct responses to famous male faces in the whole face condition and familiarity verification session (to ensure that the latency analysis is based on sufficient responses); b) produce 75% correct responses to famous male and female faces and unfamiliar faces pooled together, in the whole face condition and familiarity verification session (to prevent inclusion of data derived from response strategies); c) produce 75% correct responses to unfamiliar faces in the internal condition (to ensure that despite the high error rates to famous internal features, correct 'familiar' responses are likely to be valid and not the product of a response strategy).

The error analysis was based on a global measure which calculated errors arising for the following reasons: a) incorrect familiarity decisions made to known faces (a face is 'known' if it is recognized in the familiarity verification session); b) a failure to recognise a face in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect familiarity decisions made to known faces, 83 (6.5%) whole face responses and 320 (25.0%) internal feature responses were removed. In addition, 34 (2.7%) responses from each condition were removed as a consequence of failing to recognise faces in the familiarity verification session. As a result of eliminating outliers slower than 3 standard deviations away from the mean and faster than 300ms, 20 (1.6%) whole face and 24 (1.9%) internal feature responses were removed. Outliers were calculated for each face type separately.

Only correct familiarity decisions and those made to faces successfully recognized in the familiarity verification session, were submitted to the latency analysis. Table 7.5 shows the mean correct response latencies in each condition of the experiment in addition to associated error rates. For inspection purposes, Table 7.5 additionally reports '% misclassifications' which denotes incorrect familiarity decisions made to known faces.

Table 7.5. *Mean RTs and Percentage Errors for Familiarity Decisions in Each Condition (Experiment 9).*

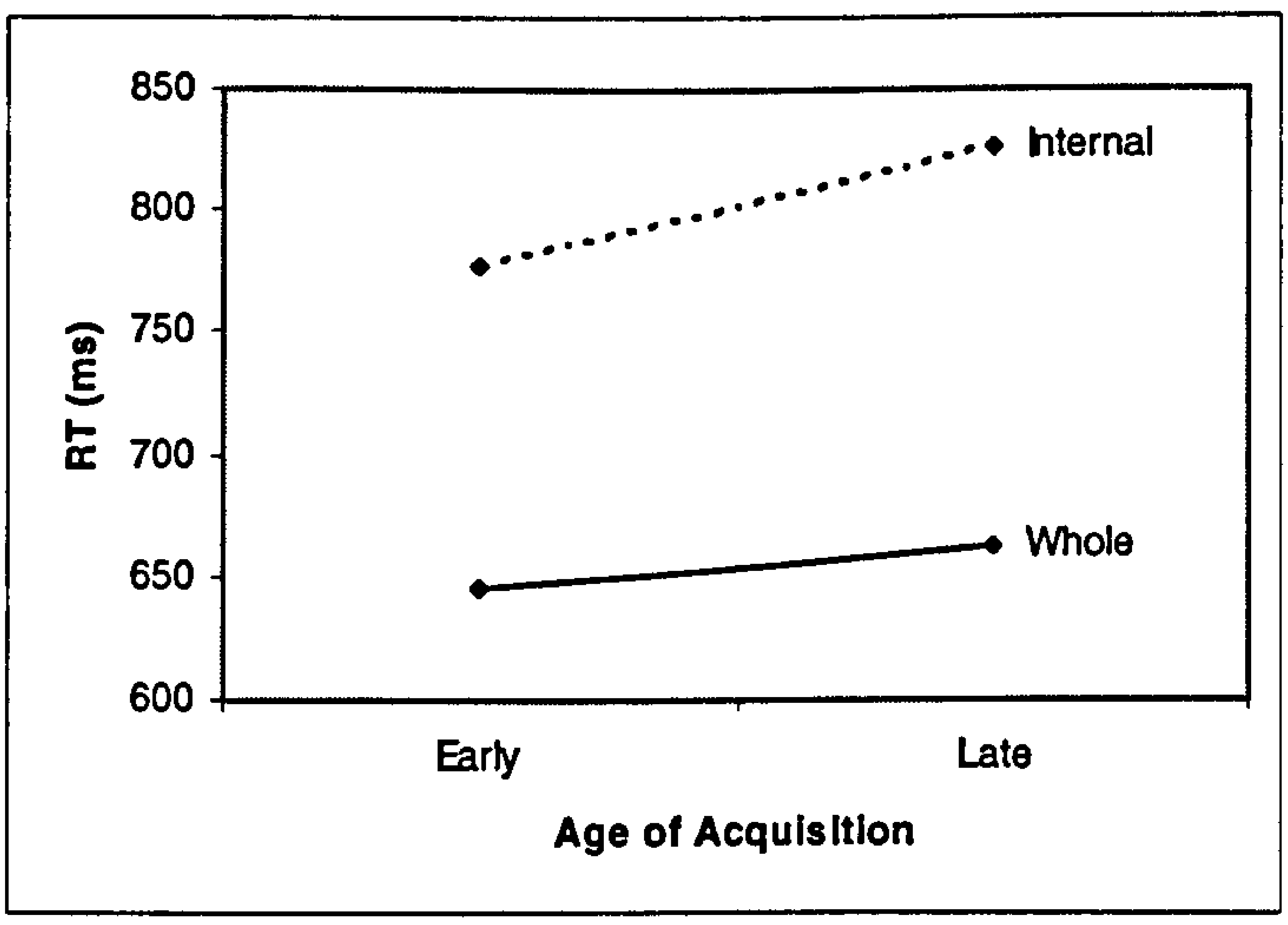
		Early Acquired		Late acquired		Unfamiliar	
Stimulus		High gender typ.	Low gender typ.	High gender typ.	Low gender typ.	High gender typ.	Low gender typ.
Whole	*Mean RT*	650	639	626	700	684	703
	SD	88	77	78	103	116	121
	% Error	9.4	7.5	8.8	17.2	6.6	8.4
	% Misclass.	4.8	5.1	5.8	11.1	2.4	2.6
Internal	*Mean RT*	754	800	780	874	878	868
	SD	168	170	166	260	238	240
	% Error	21.9	33.8	22.2	40.3	18.1	24.1
	% Misclass.	18.3	31.2	18.1	35.9	10.0	18.9

Reaction Times. Reaction times to the famous male faces were analysed by-subjects with face type (whole or internal), AoA (early or late) and gender typicality (high or low) as within-subjects factors, and by-items with face type, AoA and gender typicality as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of face type was significant in both analyses, $F_1(1,31) = 49.76$, MSe = 28,332.56, $p<.001$; $F_2(1,72) = 54.91$, MSe = 4,869.17, $p<.001$, with wholes (654 ms) recognized faster than internals (802 ms). AoA was significant by-subjects, $F_1(1,31) = 17.62$, MSe = 4,292.78, $p<.001$, but not by-items, $F_2(1,72) = 3.02$, MSe = 4,869.17, $p = .09$, with early acquired faces

(711 ms) recognized faster than late acquired faces (745 ms). The main effect of gender typicality was significant in both analyses, $F_1(1,31) = 26.59$, MSe = 6,242.42, $p<.001$; $F_2(1,72) = 4.63$, MSe = 4,869.17, $p<.05$, stemming from the faster recognition of gender typical faces (753 ms) compared with those low in gender typicality (703 ms).

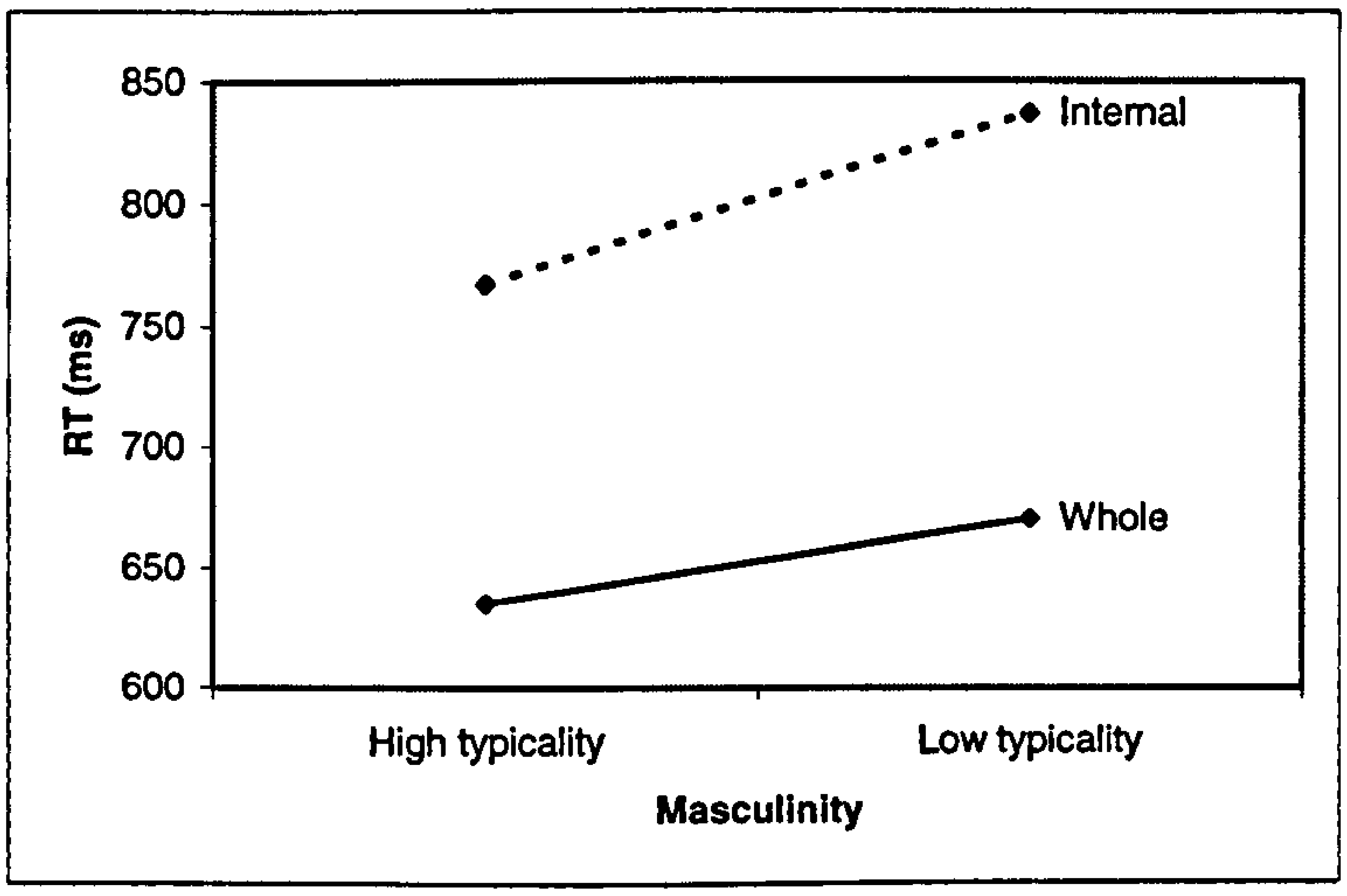
A two-way interaction between AoA and face type approached significance in the by-subjects analysis, $F_1(1,31) = 4.01$, MSe = 4,014.06, $p = .054$, see Figure 7.2. With alpha set at .05, *a posteriori* Tukey comparisons (HSD = 43) demonstrated that significant AoA effects were restricted to the recognition of internal features, whereas differences in performance as a function of face type were significant for early acquired and late acquired faces.

Figure 7.2. *Effect of AoA on Familiarity Decision RTs as a Function of Face Type (Experiment 9).*



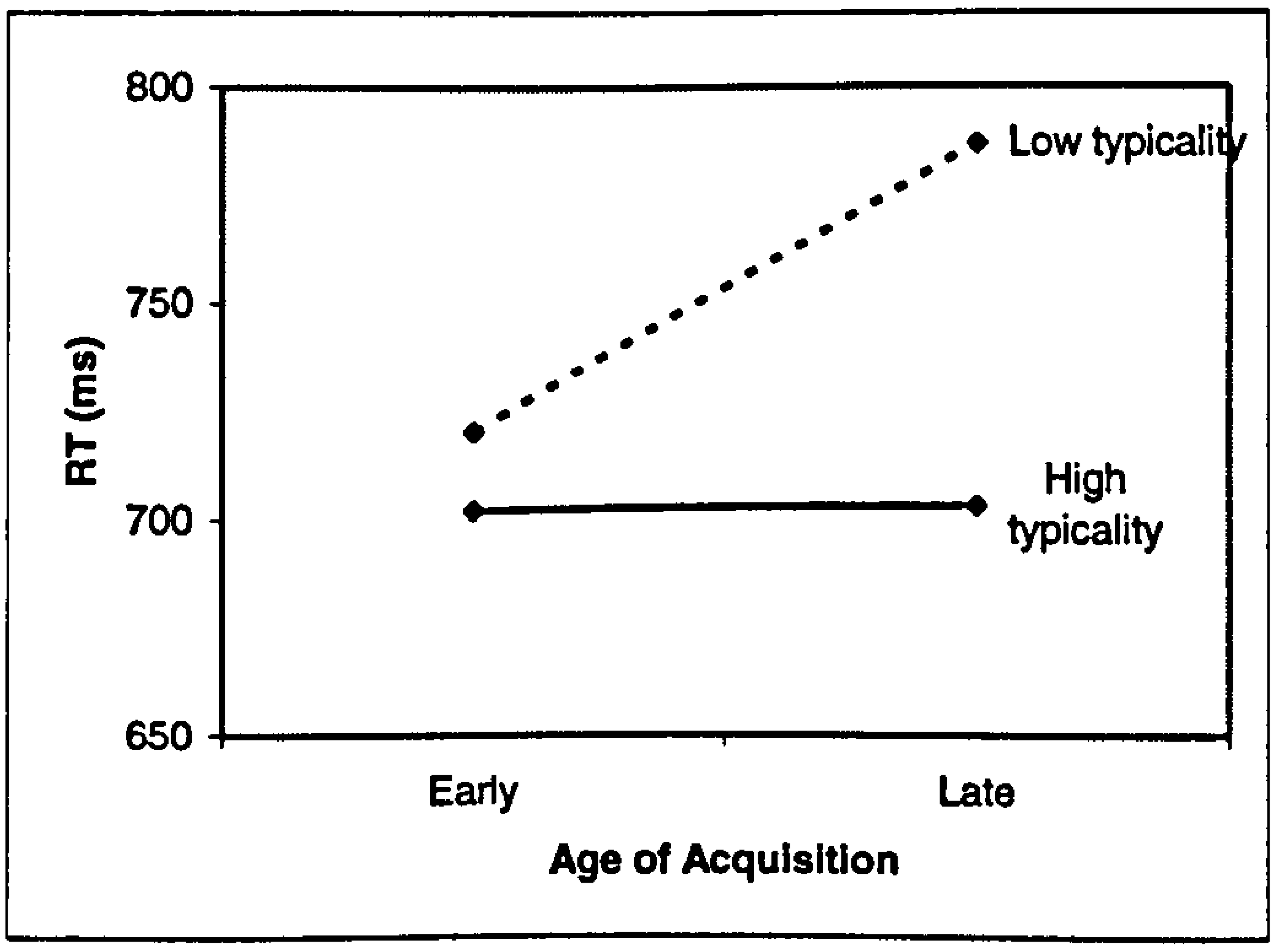
A significant two-way interaction between face type and gender typicality also emerged from the by-subjects analysis, $F_1(1,31) = 5.29$, MSe = 4,571.26, $p<.05$, see Figure 7.3. Tukey tests (HSD = 46) revealed that significant gender typicality effects were restricted to internal feature responses.

Figure 7.3. *Effect of Gender Typicality on Familiarity Decision RTs as a Function of Face Type (Experiment 9).*



The final interaction to emerge was a two-way interaction between AoA and gender typicality in the by-subjects analysis, $F_1(1,31) = 8.51$, MSe = 8,155.04, $p<.01$, see Figure 7.4. Tukey tests (HSD = 61) demonstrated that AoA exclusively affected recognition performance associated with faces low in gender typicality, and effects of gender typicality were restricted to the recognition of late acquired faces. No further interactions emerged by-subjects or by-items.

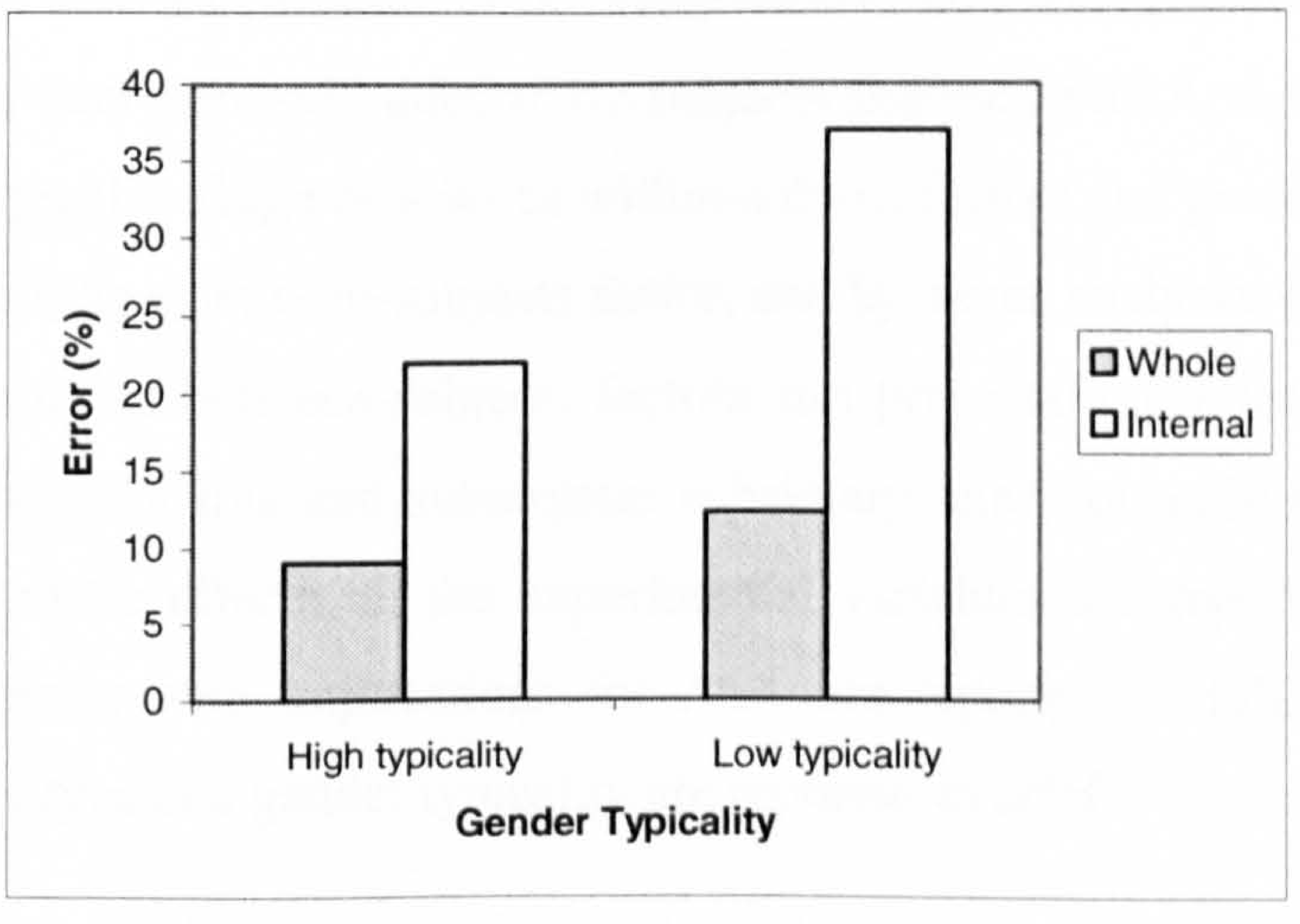
Figure 7.4. *Effect of AoA on Familiarity Decision RTs as a Function of Gender Typicality (Experiment 9).*



Errors. Errors made to the famous male faces were analysed by-subjects with face type (whole or internal), AoA (early or late) and gender typicality (high or low) as within-subject factors, and by-items with face type, AoA and gender typicality as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(1,31) = 35.39$, MSe = 6.41, $p<.001$; $F_2(1,72) = 30.86$, MSe = 23.53, $p<.001$, confirming the observed tendency for wholes (mean n = 1.1, 10.7%) to be recognized more accurately than internals (mean n = 3.0, 29.5%). The main effect of AoA was significant by-subjects, $F_1(1,31) = 12.87$, MSe = 0.79, $p = .001$, but not by-items, $F_2(1,72) = 1.38$, MSe = 23.53, $p = .24$, with early acquired faces (mean n = 1.8, 18.1%) recognised more accurately than late acquired faces (mean n = 2.2, 22.1%). The main effect of gender typicality was significant in both analyses, $F_1(1,31) = 38.86$, MSe = 1.38, $p<.001$; $F_2(1,72) = 7.27$, MSe = 23.53, $p<.01$, with gender typical faces (mean n = 1.6, 15.6%) recognised more accurately than faces low in gender typicality (mean n = 2.5, 24.7%).

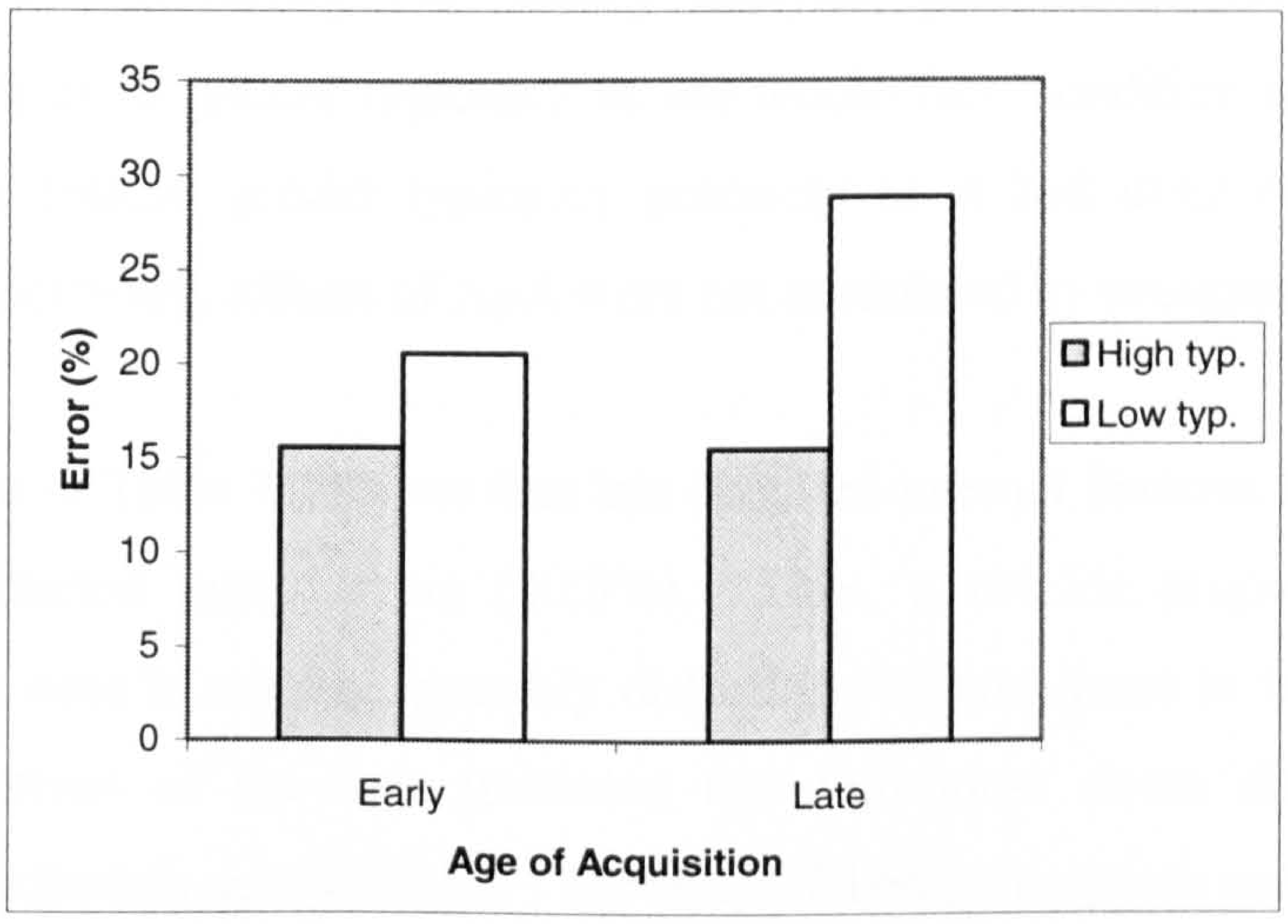
A two-way interaction between face type and gender typicality was significant in the by-subjects analysis, $F_1(1,31) = 18.09$, MSe = 1.22, $p<.001$, see Figure 7.5. *A posteriori* Tukey comparisons (HSD = 0.8) demonstrated that gender typicality solely affected internal feature recognition accuracy, whereas effects of face type generalised across both levels of gender typicality.

Figure 7.5. *Effect of Gender Typicality on Familiarity Decision Accuracy as a Function of Face Type (Experiment 9).*



A two-way interaction between AoA and gender typicality also emerged from the by-subjects analysis, $F_1(1,31) = 13.80$, $MSe = 0.80$, $p = .001$, see Figure 7.6. Tukey tests ($HSD = 0.61$) revealed that effects of AoA were restricted to performance accuracy associated with faces low in gender typicality, and effects of gender typicality were restricted to the late acquired condition. No further interactions emerged by-subjects or by-items.

Figure 7.6. *Effect of AoA on Familiarity Decision Accuracy as a Function of Gender Typicality (Experiment 9).*



7.4.3.1 Presentation Order Analysis

A collection of subsidiary ANOVAs explored the effect of presentation order on each face type to assess whether or not systematic effects were imparted. Reaction time and error data were submitted to by-subjects analyses with AoA (early or late) and gender typicality (high or low) as within-subject factors and presentation order (first or second) as a between-subjects factor, and by-items analyses with AoA and gender typicality as between-subjects factors and presentation order as a within-subjects factor. For this and subsequent subsidiary analyses, only results which indicate systematic effects on the experimental variables are reported. Further, only those results with implications for AoA are reported in full. Systematic effects on face type and gender typicality are reported in brief.

Reaction Times. Presentation order failed to modulate effects of gender typicality in the whole face condition. However, an interaction between gender typicality and presentation order emerged from the by-subjects analysis of internal features. In short, presentation order caused reaction times to decrease significantly over successive presentations for gender typical faces and those low in gender typicality. Additionally, the effect of gender typicality emerged during the first presentation only. Importantly, effects of AoA were not modulated by presentation order.

Errors. Reflecting the latency analysis, presentation order failed to modulate effects of gender typicality in the whole face condition and interacted with internal feature gender typicality precisely as it had done in the latency analysis. Importantly, effects of AoA were not modulated by presentation order.

An inspection of Table 7.5 shows that late acquired internal features low in gender typicality attracted many errors (40.3%). Thus, a notable proportion of this reaction time data is missing, possibly distorting reaction times in this condition. Closer inspection of the data indicated that that most errors stemmed from responses to internals seen in block 1 (55.0%). When participants saw whole faces before internals, internal features were recognised more accurately (25.6%).

As the presentation order analysis had demonstrated order effects on gender typicality in the internal condition, it seemed conceivable that gender typicality effects could be artefacts associated with the high error rates. Therefore, the latency data were reanalysed with the block 1 data removed. A mixed design ANOVA was carried out with face type as a between-subjects factor[23] and AoA and gender typicality as within-subjects factors. Face type was not significant, $F_1(1,30) = 0.53$, MSe = 36,118.47, $p = .47$. However, AoA, $F_1(1,30) = 12.53$, MSe = 2,190.62, $p = .001$, and gender typicality, $F_1(1,30) = 9.85$, MSe = 4,492.26, $p<.01$, reached significance, preserving effects from the original analysis. A three-way interaction comprising face type, AoA and gender typicality emerged from the analysis, $F_1(1,30) = 7.16$, MSe = 2,467.90, $p = .01$. ANOVAs with AoA and

[23] The participants who saw whole faces in block 2 were different to those who saw internal features in block 2, thus when order is taken into account, face type becomes a between-subjects factor.

gender typicality as factors were carried out to scrutinise effects for each face type separately. For whole faces, AoA, $F_1(1,15) = 4.68$, MSe = 2,897.85, $p<.05$, and gender typicality, $F_1(1,15) = 10.31$, MSe = 2,482.87, $p<.01$, were significant and interacted significantly, $F_1(1,15) = 12.00$, MSe = 2,113.70, $p<.01$. Tukey tests (HSD = 47) showed that AoA effects were restricted to faces low in gender typicality and effects of gender typicality were restricted to late acquired faces. For internal features, AoA was significant, $F_1(1,15) = 9.35$, MSe = 1,483.40, $p<.01$, but gender typicality, $F_1(1,15) = 2.91$, MSe = 6,501.65, $p = .11$, and the interaction, $F_1(1,15) = 0.29$, MSe = 2,822.10, $p = .60$, were not.

Compared with the original full analysis, the extent of the AoA effects did not change. However, gender typicality effects became less prevalent, affecting late acquired whole faces only. This reduced effect does not constitute compelling evidence in favour of an effect of gender on identity processing.

7.4.3.2 Familiarity Analysis

Post-hoc analyses were carried out on the famous face and unfamiliar face data to assess the impact of familiarity on familiarity decisions. These analyses were carried out to provide a comparison with post-hoc familiarity analyses of gender classification performance (Experiment 10).

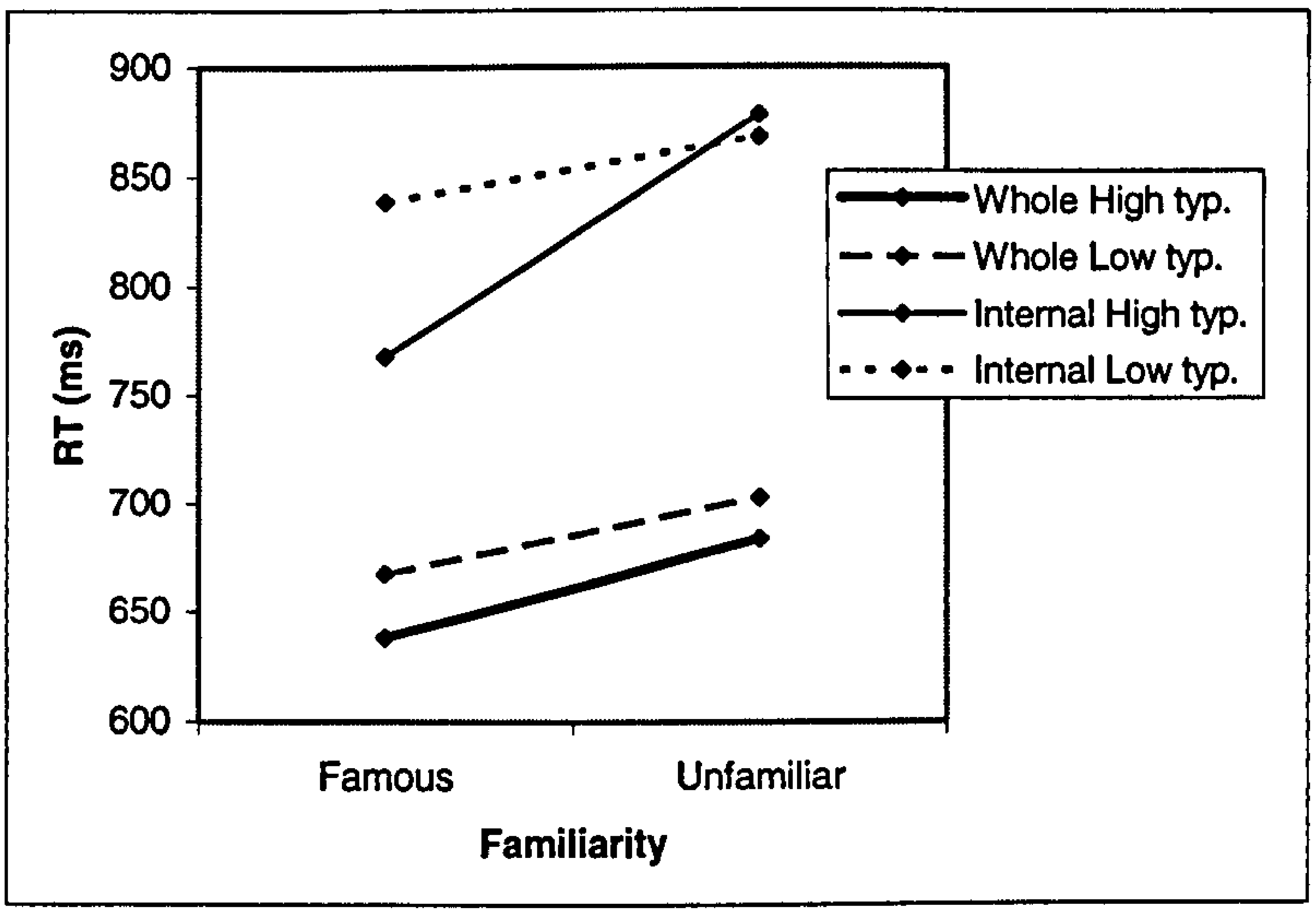
For the unfamiliar faces, the error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect familiarity decisions made to faces correctly identified as 'unfamiliar' in the familiarity verification session; b) a failure to correctly identify a face as 'unfamiliar' in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect familiarity decisions, 31 (2.4%) whole face responses and 79 (6.2%) internal feature responses were removed. In addition, 40 (3.1%) responses from each face type were removed for failing to recognise faces in the familiarity verification session. Further, 25 (2.0%) whole face responses were deemed outliers, as were 22 (1.7%) internal feature responses.

Only correct familiarity decisions, and those made to faces correctly classified as 'unfamiliar' in the familiarity verification session, were submitted to the latency analysis. See Table 7.5 for mean famous and unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects with face type (whole or internal), familiarity (famous or unfamiliar) and gender typicality (high or low) as within-subjects factors, and by-items with face type, familiarity and gender typicality as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(1,31) = 54.41$, MSe = 31,915.79, $p<.001$; $F_2(1,152) = 203.86$, MSe = 3,862.25, $p<.001$, with wholes (673 ms) recognized more quickly than internals (838 ms). The main effect of familiarity was also significant in both analyses, $F_1(1,31) = 13.32$, MSe = 14,452.09, $p = .001$; $F_2(1,152) = 27.95$, MSe = 3,862.25, $p<.001$, with famous faces (728 ms) recognised faster than unfamiliar faces (783 ms). The main effect of gender typicality was significant by-subjects, $F_1(1,31) = 12.20$, MSe = 3,924.73, $p = .001$ but not by-items, $F_2(1,152) = 3.38$, MSe = 3,862.25, $p = .07$, with gender typical faces (742 ms) recognised faster than faces low in gender typicality (769 ms).

A two-way interaction between familiarity and gender typicality emerged from the by-subjects analysis, $F_1(1,31) = 9.96$, MSe = 3,508.03, $p<.01$, and a two-way interaction between familiarity and face type emerged by-items, $F_2(1,152) = 6.31$, MSe = 3,862.25, $p = .01$. Moreover, a three-way interaction incorporating familiarity, gender typicality and face type emerged from the by-subjects analysis, $F_1(1,31) = 4.95$, MSe = 3,919.07, $p<.05$, see Figure 7.7. As a means of clarifying the nature of the familiarity effect, data from each face type were submitted independently to 2 x 2 within-subjects ANOVAs with familiarity and gender typicality as factors.

Figure 7.7. *Relationship Between Face Type and Gender Typicality as a Function of Familiarity, Indexed by Familiarity Decision RTs (Experiment 9).*

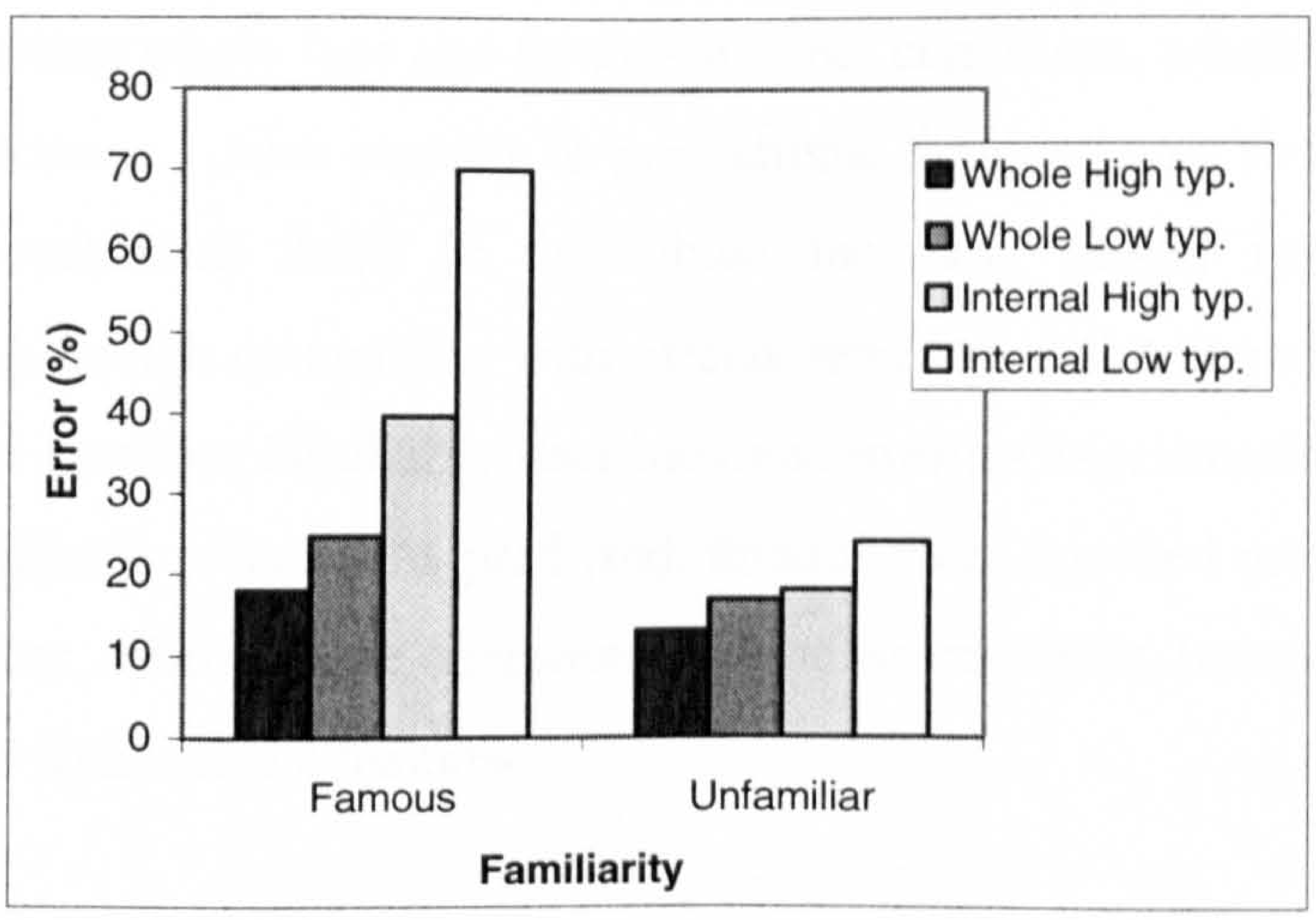


In the whole face condition, the main effects of familiarity, $F_1(1,31) = 12.43$, MSe = 4,201.56, p = .001, and gender typicality, $F_1(1,31) = 22.57$, MSe = 815.59, p<.001, were significant. The interaction was not significant, $F_1(1,31) = 0.71$, MSe = 1,600.36, p = .41. For internal features, familiarity, $F_1(1,31) = 10.86$, MSe = 14,138.55, p<.01, and gender typicality, $F_1(1,31) = 4.61$, MSe = 6,557.06, p<.05, were significant, as was the interaction, $F_1(1,31) = 9.13$, MSe = 5,826.74, p<.01. Tukey tests (α = .05) demonstrated that familiarity effects were restricted to the recognition of gender typical faces, and effects of gender typicality exclusively modulated famous face recognition (HSD = 52).

Errors. The data were analysed by-subjects with face type (whole or internal), familiarity (famous or unfamiliar) and gender typicality (high or low) as within-subjects factors, and by-items with face type, familiarity and gender typicality as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(1,31) = 28.28$, MSe = 8.84, p<.001; $F_2(1,152) = 36.28$, MSe = 14.09, p<.001, with wholes (mean n = 1.8, 9.1%) recognised more accurately than internal features (mean n = 3.8, 19.0%). The main effect of familiarity was significant in both analyses, $F_1(1,31) = 18.19$, MSe = 14.18, p<.001; $F_2(1,152) = 34.28$, MSe = 14.09, p<.001, with famous faces (mean n = 3.8, 19.1%) recognised less accurately than unfamiliar faces (mean n = 1.8, 9.1%). In addition, the main effect of gender typicality was significant in both analyses, $F_1(1,31) = 31.12$, MSe $= 2.79$, $p<.001$; $F_2(1,152) = 6.18$, MSe $= 14.09$, $p<.05$, with fewer errors made to gender typical faces (mean n = 2.2, 11.2%) compared with faces low in gender typicality (mean n = 3.4, 17.0%).

In by-subjects and by-items analyses, significant two-way interactions emerged between familiarity and face type, $F_1(1,31) = 18.45$, MSe $= 6.48$, $p<.001$; $F_2(1,152) = 17.04$, MSe $= 14.09$, $p<.001$, and between familiarity and gender typicality, $F_1(1,31) = 10.65$, MSe $= 2.78$, $p<.01$; $F_2(1,152) = 5.97$, MSe $= 14.09$, $p<.05$. Further, the interaction between face type and gender typicality was significant by-subjects, $F_1(1,31) = 12.05$, MSe $= 2.23$, $p<.01$, and approached significance by-items, $F_2(1,152) = 3.44$, MSe $= 14.09$, $p = .07$. By-subjects, these interactions were subsumed in a three-way interaction between familiarity, gender typicality and face type, $F_1(1,31) = 9.94$, MSe $= 1.87$, $p<.01$, see Figure 7.8. As a means of clarifying the nature of the familiarity effect, data from each face type were submitted independently to 2 x 2 within-subjects ANOVAs with familiarity and gender typicality as factors.

Figure 7.8. *Relationship Between Face Type and Gender Typicality as a Function of Familiarity, Indexed by Familiarity Decision Accuracy (Experiment 9).*



In the whole face condition, whilst gender typicality was significant, $F_1(1,31) = 6.72$, MSe $= 1.27$, $p<.05$, the main effect of familiarity, $F_1(1,31) = 2.74$, MSe $= 4.79$, $p = .11$, and the interaction between familiarity and gender typicality was not significant, $F_1(1,31) = 0.67$, MSe = 0.94, p = .42. For internal features, familiarity, $F_1(1,31) = 22.97$, MSe = 15.87, p<.001, and gender typicality, $F_1(1,31) = 28.00$, MSe = 3.75, p<.001, were significant, as was the interaction, $F_1(1,31) = 12.82$, MSe = 3.71, p<.001. Tukey tests (HSD = 1.3) demonstrated that familiarity affected gender typical faces and those low in gender typicality, and effects of gender typicality exclusively modulated famous face recognition.

7.4.4 Discussion

Whilst AoA did affect response latencies as predicted, effects were restricted to internal features and faces low in gender typicality. Nonetheless, there was a tendency for early acquired whole faces (645 ms) to be recognised faster than late acquired whole faces (663 ms). However, there were no trends for gender typical items. Contrary to predictions, gender typicality exerted effects, modulating performance with internal features and late acquired items. An analysis of the errors revealed an almost identical pattern of results, except AoA effects ensued regardless of face type: speed-accuracy trade-offs appeared not to compromise the results.

Contrary to predictions, gender typicality exerted effects in the post-hoc familiarity analysis, affecting whole face and famous internal conditions, whether indexed by latency or accuracy. Also counter to predictions, famous items were recognised faster than unfamiliar items in the whole face and gender typical internal conditions. Although compatible with effects reported in the literature (Bruce et al., 1987a) the previous familiarity decision experiments (Experiments 5 and 7) did not yield familiarity effects. As predicted, famous faces incurred more errors than unfamiliar faces, reflecting the operation of response strategies, however this effect only occurred with internal features.

7.5 Experiment 10

7.5.1 Introduction

Experiment 10 repeated Experiment 9 but as a gender classification task. Gender decisions were made to whole faces and internal features manipulated factorially according to AoA and gender typicality. Once again experimental items constituted male faces with female faces employed as filler items. A familiarity verification session was administered after the experiment proper.

Consistent with parallel-route accounts which suggest that identity information has the opportunity to influence gender decisions whenever gender processing is sufficiently slowed, it was hypothesised that AoA would modulate gender decisions made to faces low in gender typicality. This effect was predicted for whole faces and internal features. Gender typical faces were only predicted to elicit AoA effects in the internal condition; where external gender cues had been removed. However, it was anticipated that this AoA effect would be subdued compared to effects obtained with low gender typicality internals. In addition, post-hoc familiarity effects were predicted to reflect the AoA effects: famous faces were predicted to be classified faster and more accurately than unfamiliar faces in the low gender typicality conditions and also in the gender typical internal features condition.

7.5.2 Method

Participants. Thirty two undergraduate students from the University of York took part in this experiment (8 male, 24 female), receiving course credit or a payment of £2 for their participation. Participants were 19 – 21 years old (mean = 19.97; S.D. = 0.74) and were required to have lived in the UK for at least 18 years.

Design. The experimental variables comprised face type (with two levels; whole face or internal features), AoA (with two levels; early or late) and gender typicality (with two levels; high or low) which were explored in a repeated measures design. The order of the face types was counterbalanced across participants.

The dependent variables, response latency and accuracy, were the products of a gender decision task.

Materials. See Experiment 9 details, as the same stimuli were employed, p. 218.

Procedure and Apparatus. See Experiment 9 details, p. 222. Note, however, that during the experiment participants were instructed to "decide as quickly and accurately as possible, whether or not the presented face is male or female".

7.5.3 Results

Four participants had to be replaced after failing to meet the criteria which required them to: a) produce 75% correct responses to famous male faces in the whole face condition and familiarity verification session; b) produce 75% correct responses to famous male and female faces and unfamiliar faces pooled together, in the whole face condition and familiarity verification session; c) produce 75% correct responses to unfamiliar faces in the internal condition.

The error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications made to known faces (a face is 'known' if it is successfully recognized in the familiarity verification session); b) a failure to recognise a face in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect gender classifications made to known faces, 49 (3.8%) whole face responses and 51 (4.0%) internal feature responses were removed. In addition, 93 (7.3%) responses from both face type conditions were removed for failing to recognise faces in the familiarity verification session. Outliers slower than 3 standard deviations away from the mean and faster than 300ms were eliminated. As a consequence, 45 (3.5%) whole face and 24 (1.9%) internal feature responses were removed. Outliers were calculated for each face type separately.

Only correct gender classifications, and those made to faces successfully recognized in the familiarity verification session, were submitted to the latency analysis. Table 7.6 shows the mean correct response latencies in each condition of the experiment in addition to associated error rates. For inspection purposes, Table 7.6 additionally reports '% misclassifications' which denotes incorrect gender classifications made to faces classified correctly during the familiarity verification session.

Table 7.6. *Mean RTs and Percentage Errors for Gender Decisions in Each Condition (Experiment 10).*

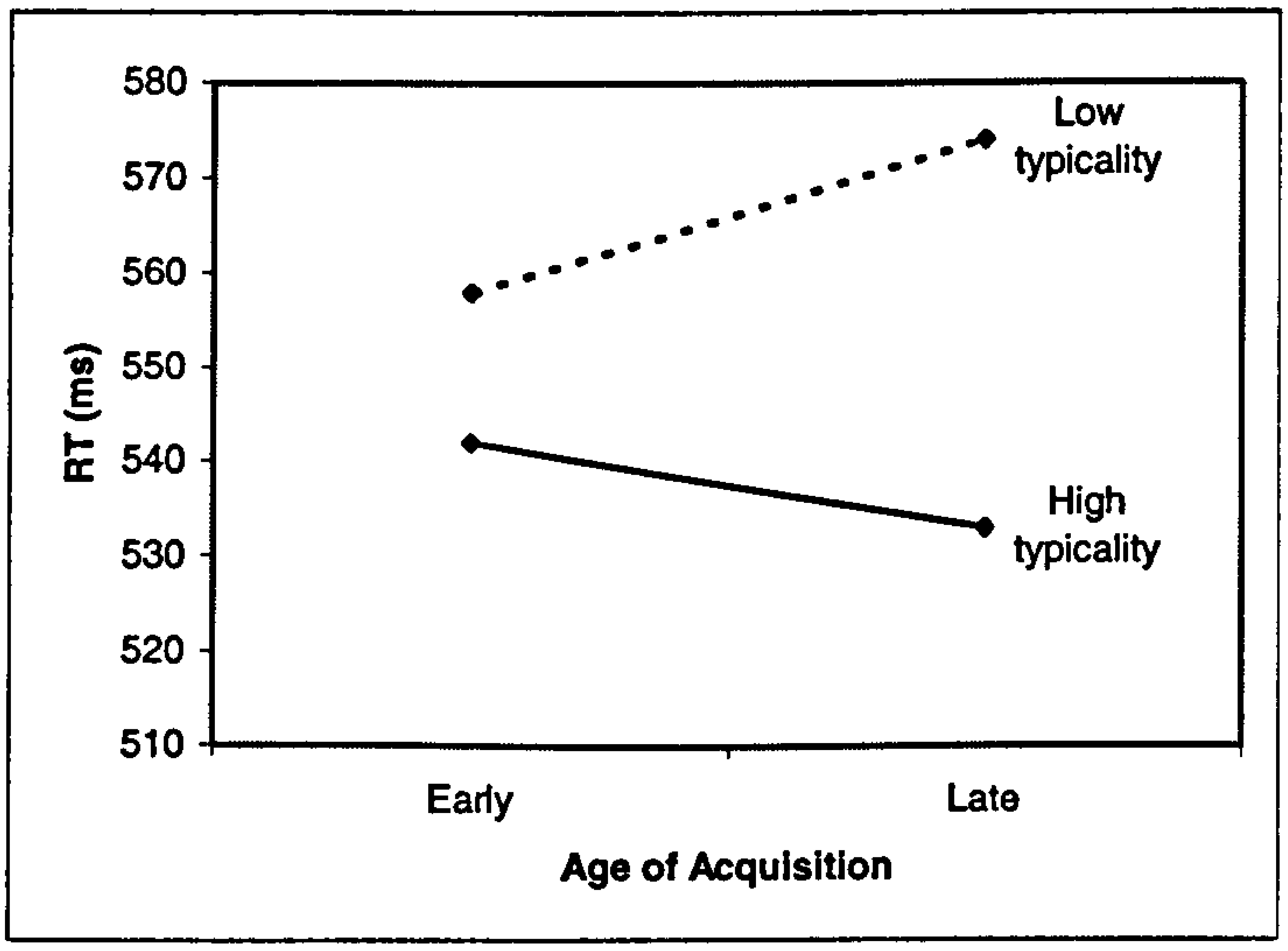
		Early Acquired		Late acquired		Unfamiliar	
Stimulus		Gender typ.	Low gender typ.	Gender typ.	Low gender typ.	Gender typ.	Low gender typ.
Whole	*Mean RT*	514	530	496	539	485	500
	SD	90	92	82	90	67	75
	% Error	10.6	15.0	12.2	20.6	8.8	12.5
	% Misclass.	1.6	4.6	3.3	7.2	2.6	3.0
Internal	*Mean RT*	570	586	570	606	558	586
	SD	91	100	95	101	77	98
	% Error	9.4	14.4	10.0	18.8	6.7	12.0
	% Misclass.	2.7	7.7	2.3	4.5	3.0	3.8

Reaction Times. Reaction times to the famous male faces were analysed by-subjects with face type (whole or internal), AoA (early or late) and gender typicality (high or low) as within-subjects factors, and by-items with face type, AoA and gender typicality as between-subjects factors. For this and subsequent analyses, by-subject data is graphed and submitted to post-hoc tests and reported means stem from this analysis. The main effect of face type was significant in both analyses, $F_1(1,31) = 32.50$, MSe = 8,014.54, $p<.001$; $F_2(1,72) = 31.08$, MSe = 2,623.15, $p<.001$, with whole faces (520 ms) classified according to gender faster than internal features (584 ms). The main effect of gender typicality was also significant in both analyses, $F_1(1,31) = 19.64$, MSe = 2,654.96, $p<.001$; $F_2(1,72) =$

6.41, MSe = 2,623.15, $p<.05$, with gender typical faces (537 ms) classified faster than faces low in gender typicality (566 ms). The main effect of AoA was not significant in either analysis, $F_1(1,31) = 0.37$, MSe = 2,115.02, p = .55; $F_2(1,72) =$ 0.11, MSe = 2,623.15, p = .74.

A two-way interaction between AoA and gender typicality approached significance by-subjects, $F_1(1,31) = 4.13$, MSe = 2,191.06, p = 051, see Figure 7.9. With alpha set at .05, Tukey tests (HSD = 32) demonstrated that AoA failed to exert significant effects regardless of gender typicality. The significant effect of gender typicality was restricted to late acquired faces.

Figure 7.9. *Effect of AoA on Gender Decision RTs as a Function of Gender Typicality (Experiment 10).*



Errors. Errors made to famous male faces were analysed by-subjects with face type (whole or internal), AoA (early or late) and gender typicality (high or low) as within-subject factors, and by-items with face type, AoA and gender typicality as between-subjects factors. The main effects of face type, $F_1(1,31) =$ 2.85, MSe = 1.00, p = .10; $F_2(1,72) = 0.96$, MSe = 8.82, p = .33, and AoA were not significant in either analysis, $F_1(1,31) = 3.10$, MSe = 1.92, p = .09; $F_2(1,72) = 2.05$, MSe = 8.82, p = .16. Gender typicality was significant in both analyses, $F_1(1,31) =$ 11.48, MSe = 2.57, $p<.01$; $F_2(1,72) = 10.48$, MSe = 8.82, $p<.01$, with gender typical faces (mean n = 1.0, 10.2%) classified according to gender more accurately than faces low in gender typicality (mean n = 1.70, 17.0%). No significant interactions emerged from either analysis.

7.5.3.1 Presentation Order Analysis

A collection of subsidiary ANOVAs explored the effect of presentation order on each face type to assess whether or not systematic effects were imparted. Reaction time and error data were submitted to by-subjects analyses with AoA (early or late) and gender typicality (high or low) as within-subject factors and presentation order (first or second) as a between-subjects factor, and by-items analyses with AoA and gender typicality as between-subjects factors and presentation order as a within-subjects factor.

Reaction Times. Presentation order failed to modulate effects of AoA or gender typicality in the whole face and internal feature conditions.

Errors. Presentation order failed to modulate effects of AoA or gender typicality in the whole face and internal feature conditions.

7.5.3.2 Familiarity Analysis

Famous face and unfamiliar face data were analysed together to provide an additional tool with which to assess the notion that gender classifications are carried out independently of identity information. This analysis also provided an additional opportunity to monitor the impact of gender typicality on this relationship.

For the unfamiliar faces, the error analysis was based on a global measure which calculates errors arising for the following reasons: a) incorrect gender classifications made to faces correctly identified as 'unfamiliar' in the familiarity verification session; b) a failure to correctly identify a face as 'unfamiliar' in the familiarity verification session; c) reaction times deemed outliers. As a result of incorrect gender classifications, 35 (2.7%) whole face responses and 43 (3.4%) internal feature responses were removed. In addition, 54 (4.2%) responses from each face type condition were removed as a consequence of failing to recognise faces in the familiarity verification session. Further, 47 (3.7%) responses from the whole condition were deemed outliers, as were 23 (1.8%) internal condition responses.

Only correct gender classifications and those made to faces correctly classified as 'unfamiliar' in the familiarity verification session, were submitted to the latency analysis. See Table 7.6 for mean unfamiliar face response latencies and error rates.

Reaction Times. The data were analysed by-subjects with face type (whole or internal), familiarity (famous or unfamiliar) and gender typicality (high or low) as within-subjects factors, and by-items with face type, familiarity and gender typicality as between-subjects factors. The main effect of face type was significant in both analyses, $F_1(1,31) = 51.22$, MSe = 6,449.22, $p<.001$; $F_2(1,152) = 9.38$, MSe = 1,718.75, $p<.01$, with whole faces (506 ms) classified according to gender faster than internal features (578 ms). The main effect of familiarity was also significant in both analyses, $F_1(1,31) = 28.80$, MSe = 1,029.15, $p<.001$; $F_2(1,152) = 9.38$, MSe = 1,718.75, $p<.01$, with famous faces (552 ms) classified more slowly than unfamiliar faces (532 ms). The main effect of gender typicality was significant in both analyses, $F_1(1,31) = 31.30$, MSe = 1,266.81, $p<.001$; $F_2(1,152) = 13.83$, MSe = 1,718.75, $p<.001$, with gender typical faces (530 ms) classified according to gender faster than faces low in gender typicality (555 ms). No interactions stemmed from the analyses.

Errors. The data were analysed by-subjects with face type (whole or internal), familiarity (famous or unfamiliar) and gender typicality (high or low) as within-subjects factors, and by-items with face type, familiarity and gender typicality as between-subjects factors. The main effect of face type was not significant in either analysis, $F_1(1,31) = 2.40$, MSe = 2.11, $p = .13$; $F_2(1,152) = 1.82$, MSe = 6.06, $p = .18$. The main effect of familiarity approached significance by-subjects, $F_1(1,31) = 3.58$, MSe = 7.33, $p = .068$, and attained significance by-items, $F_2(1,152) = 8.73$, MSe = 6.06, $p<.01$, with famous faces (mean n = 2.6, 13.2%) attracting more errors than unfamiliar faces (mean n = 2.0, 10.0%). Further, the main effect of gender typicality was significant in both analyses,

$F_1(1,31) = 18.26$, MSe = 4.44, $p<.001$; $F_2(1,152) = 21.38$, MSe = 6.06, $p<.001$, with gender typical faces (mean n = 1.8, 8.8%) attracting fewer errors than faces low in gender typicality (mean n = 2.9, 14.4%). No interactions were significant in either analysis.

7.5.4 Discussion

AoA failed to exert significant effects on gender decision latencies. An interaction between AoA and gender typicality reached borderline significance (p = .051) reflecting an effect of gender typicality on late acquired faces. There was a trend for early acquired items (558 ms) to be classified according to gender faster than late acquired items (574 ms) in the low gender typicality condition. Not only was this effect noticeably weaker in the gender typical condition, thus approximating predictions, but late acquired items (533 ms) were classified marginally faster than early acquired items (542 ms). Effects of AoA were not significant in the accuracy analysis, indicating that trends in the latency analysis were not likely to be the product of speed-accuracy trade-offs. Analyses exploring presentation order revealed no systematic effects.

These trends are not entirely compatible with predictions. Although AoA was predicted to affect low gender typicality items, AoA was also predicted to affect gender typical internals: the removal of external cues to gender should be sufficient to invoke an effect of identity on gender processing. Inspection of reaction times to gender typical internal features confirmed that there were no trends for early acquired items (570 ms) to be classified faster than late acquired items (570 ms). Thus, effects of AoA on gender processing appear to be modulated solely by gender typicality, not face type.

An inspection of the mean ratings for the stimulus sets (Table 7.3) sought to determine whether or not the borderline interaction between AoA and gender typicality could be an artefact of imbalanced matching. Given that the AoA effects reported are merely trends, it is feasible that they could be the product of sub-optimally matched sets. Indeed, the difference between mean rated AoA for the gender typical early acquired and late acquired sets was 1.12 compared with a larger difference of 1.31 between the low gender typicality early and late sets. However, this difference is minimal and offset by other differences in the ratings that reduce the likelihood of a larger AoA effect for items low in gender typicality. For example, low gender typicality late acquired faces have a higher frequency rating (2.75) than their early acquired counterparts (2.46). On balance, there are no clear indications that poor matching is responsible for the tendency for faces low in gender typicality to show AoA effects.

Counter to predictions, famous faces were classified according to gender more slowly and less accurately than unfamiliar faces, despite matching gender typicality ratings across sets. Limited resources meant that the famous and unfamiliar sets were not optimally matched: in the whole face condition, unfamiliar faces had a higher mean gender typicality rating (4.28) than famous faces (4.07). However, on a 7-point scale, a difference of a 5^{th} of a rating is not particularly large. Also, famous internal features had a higher rating (4.25) than unfamiliar internals (4.13). It could be argued that if resources prohibit stringent matching, it is wise to rely on random variation in gender typicality rather than attempting to match sets. However, recall the post-hoc analysis of Experiment 6 in which gender typicality was inserted as a post-hoc factor (p. 212). Gender typicality ratings were significantly higher for famous faces low in gender typicality compared with their unfamiliar counterparts, but ratings did not differ for gender typical famous and unfamiliar faces. Although this did not adversely affect Experiment 6 (because gender typicality was not explored), had these imbalances occurred presently, genuine interactions between familiarity and gender typicality would have been indistinguishable from unwanted variations in rated gender typicality.

7.6 General Discussion: Experiments 9 and 10

Experiments 9 and 10 sought to pit the single-route and parallel-route accounts against one another by investigating the impact of gender typicality on the relationship between identity and gender processing. The post-hoc analysis of Experiment 6 (p. 212), which inserted gender typicality into the experimental design, unearthed several tendencies for gender typicality to modulate this relationship. Unfortunately, Experiment 10 failed to tease out significant results, despite being specifically designed to do so. Both Experiments 9 and 10 were characterised by weak and inconsistent findings which alluded to parallel-route mechanisms rather than offering robust evidence. For the sake of brevity, accuracy analyses are not subject to discussion. These analyses largely reflected trends in the latency analysis and indicated that speed-accuracy trade-offs were not operating to compromise AoA effects.

Experiment 9 was carried out to ensure that the stimulus set readily yielded AoA effects in a task where they were expected. AoA effects were reasonably comprehensive, affecting familiarity decisions made to internal features and items low in gender typicality. There were also trends in the whole face condition. Gender typicality exerted effects on internal features and late acquired items, inadvertently providing possible evidence for an effect of gender on identity processing. Note that AoA and gender typicality failed to affect performance in the conditions with the fastest response latencies, indicating that ceiling effects could be precluding these effects. However, with regards to gender typicality, it became apparent these inconsistent effects could stem from distortions in the latency analysis associated with high error rates in certain conditions. An analysis of familiarity decisions made to items seen in block 2, carried out to by-pass the high error rates associated with block 1, revealed that effects of gender typicality on internal features were no longer significant. Therefore, in this reduced analysis AoA retained its reasonably robust influence but gender typicality effects largely dropped out.

The results of Experiments 9 are not supported by findings from the post-hoc familiarity analysis, rendering a neat interpretation of the results difficult: the effects of gender typicality on familiarity decisions are particularly prominent in the post-hoc familiarity analysis and may indicate that gender processing can facilitate recognition. However, studies typically find no evidence for this facilitation (Bruce et al., 1987a; Wild et al., 2000) and as Clutterbuck and Johnston (2004) argued, it is hard to comprehend how gender could ever be helpful in determining identity. Baudouin and Tiberghien (2002) found, however, that participants searching for a particular face rejected distractors more quickly if their gender differed from the gender of the target face. The authors hypothesised that the cognitive system enables gender information to facilitate identification, and proposed that MDS provides an interface between identity and gender processing. Experiment 4 also demonstrated that MDS seems to code gender: distinctiveness affected gender decisions even when gender typicality was inserted as a covariate into the analysis (p. 130). This finding does stray somewhat for the core objective of these experiments, namely to distinguish parallel-route and single-route accounts, but it bolsters evidence in favour of a relationship between identity and gender processing.

Experiment 10 revealed a trend for AoA to affect gender decisions made to low gender typicality items and a marginal reversed effect on gender typical items. The prediction that low gender typicality whole faces and internal features would elicit AoA effects, therefore, received some support. Contrary to predictions, AoA failed to affect gender decisions made to gender typical internals. The finding that face type failed to modulate effects of AoA on gender processing was surprising given that most studies investigating the relationship between identity and gender processing vary ease of gender processing by removing external cues. From these results, it appears that the effect of identity on gender processing is triggered exclusively by intrinsic physiognomic variations in gender typicality.

The failure to achieve significant AoA effects in Experiment 10 was disappointing and compounded by the reversed familiarity effect in the post-hoc familiarity analysis. Nevertheless, the conspicuous failure for face type to modulate the relationship between identity and gender processing provided more evidence with which to distinguish parallel-route and single-route accounts. Given that the single-route hypothesis predicts integrality when external features are removed, the failure for face type to modulate effects renders this hypothesis unlikely. On the other hand, parallel-route accounts are not devastated by the failure for face type to modulate effects. Moreover, they can accommodate the tendencies for gender typicality to modulate AoA effects.

It is possible that the gender typicality of a face does not modulate the relationship between identity and gender processing. Bruce's (1986) proposition that degree of androgyny mediates this relationship was an afterthought rather than the product of the experimental design, and studies that do report effects of identity on gender processing typically remove external cues to gender to invoke this effect. On the other hand, the striking tendencies for gender typicality to modulate AoA effects may reflect the operation of genuine mechanisms which could have been enhanced with a larger stimulus set. Indeed, the current stimulus set was less than ideal. Had the database been extended to permit reasonably sized, optimally matched sets, containing both male and female faces, significant results may have ensued. However, the database would have required considerable extension and coupled with the ongoing problem of obtaining enough stimuli sufficiently familiar to the majority of participants, this would have been a particularly time consuming undertaking. To compound these problems, the transient nature of faces means that they become decreasingly fashionable stimuli as time elapses, and decreasingly effective as a stimulus set for increasingly younger cohorts of participants.

In summary, whereas AoA affected familiarity decisions made to internal features and items low in gender typicality (Experiment 9), there were no significant effects of AoA on gender classifications (Experiment 10). However, a trend emerged which was compatible with predictions: AoA was more likely to affect gender decisions made to items low in gender typicality than gender typical items. From these results it appeared that gender typicality, rather than face type, modulated the relationship between identity and gender processing. This finding is distinctly at odds with single-route accounts and more aptly captured by parallel-route mechanisms. However, further research is needed to identify the most likely parallel-route candidate (e.g. Clutterbuck & Johnston, 2004; Rossion, 2002). Additionally, replications with an entirely different stimulus set are required to ensure that these effects are reliable.

Experiment 9 also provided evidence that gender typicality may affect identification processes, particularly in the post-hoc familiarity analysis. This is compatible with Baudouin and Tiberghien's (2002) hypothesis that MDS provides an interface between identity and gender processing whereby knowing the gender of a face can facilitate identification. Together with Experiment 10, these results indicated that the relationship between identity and gender processing could be one of mutual influence. Whether or not this relationship demands multiple mechanisms to exert bidirectional influence, is another question for future research.

CHAPTER 8 – AN OVERVIEW OF EMPIRICAL FINDINGS AND SUMMARY OF IMPLICATIONS FOR AOA AND MODELS OF FACE PROCESSING

8.1 Introduction and Summary of Main Findings

Tackled from a cognitive perspective, studies of AoA aim to clarify loci and mechanisms underpinning this effect. Since Carroll and White (1973) first reported an effect of AoA on object naming latencies, a persistent interest in the variable has shifted conception of it as a stage-dependent process to a more or less ubiquitous by-product of the way in which information is stored and accessed in the brain. This is reflected in recent implementations that refrain from attributing to AoA mechanisms specific to particular domains or task demands (A. Ellis & Lambon Ralph, 2000; Steyvers & Tenenbaum, 2005; Zevin & Seidenberg, 2002). This thesis sought to ascertain the nature of the relationship between AoA and distinctiveness in order to offer a refinement of the mechanisms underpinning AoA and concomitantly shed some light on the locus of AoA effects.

From a cognitive stance, studies of face processing seek to elucidate the different operations involved in face processing and how they interact with one another. Whereas the processes of identity and gender were historically conceived as functionally independent (Bruce and Young, 1986) single-route accounts and less radical parallel-route accounts have since pervaded the literature. This thesis recruited AoA as an investigative tool to establish the nature of the relationship between identity and gender processing and constrain mechanisms proposed to sub-serve this relationship. The mutual investigation of AoA and face processing sought to derive implications for each area of research.

Chapters 2 and 3 investigated the hypothesis that AoA and distinctiveness interact, the rationale being that the temporal dimension of distinctiveness may be entrenched in the representations of faces acquired at different points in time. Couched in terms of A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis and Valentine's (1991) MDS, interactivity between the variables would indicate that mechanisms underpinning each of the theories required elaboration.

Somewhat deceptively, the Pilot Experiment yielded the predicted interaction. This indicated that the childhood attenuation of distinctiveness effects (H. Ellis, 1992; Johnston & H. Ellis, 1995) and its emergence with experience, was manifested in the respective representations of early and late acquired faces. However, age of image had not been standardised and so it was possible that failure to standardise these images had influenced the results. Carried out with up-to-date images, Experiment 1 yielded an additive relationship between the variables. However, after eliminating poorly recognised faces a post-hoc analysis recovered the predicted interaction. Using a stimulus set screened to ensure familiarity, Experiment 2 yielded an interaction which countered predictions. In light of research which suggests that faces become increasingly distinctive with age (Deffenbacher et al., 1998; O'Toole et al., 1997), Experiment 3 was carried out with dated images. This was considered a necessary precaution in order to tap the distinctiveness of faces at the time of acquisition. Nevertheless, these results signalled additivity despite testing 49 participants.

Whilst AoA and distinctiveness emerged as robust effects, the interaction was not consistent. Inclinations towards additivity were considered to be more persuasive than various interactions recovered under suboptimal circumstances. Therefore, Chapters 2 and 3 concluded that AoA and distinctiveness are sub-served by independent mechanisms and theories of AoA and distinctiveness require no integration or elaboration. Moreover, by applying Sternberg's (1969) additive-factors logic, it was tentatively deduced that mechanisms underpinning AoA and distinctiveness could be divided into serially-arranged stages. Moreover, because distinctiveness effects are associated with FRUs (Burton et al., 1990; Newell et al., 1999) it was logically inferred that AoA must exert effects at the level of PINs, or the FRU-PIN connections, in order to affect familiarity decisions.

Chapter 4 investigated the effects of AoA and distinctiveness on gender processing to provide evidence that identity and gender processing may be interfaced in some way. Experiment 4 demonstrated that gender decision performance was affected by distinctiveness but not AoA. In view of evidence which suggests that gender decisions can be made on the basis of visually-derived semantic information without invoking identity information (Bruce, 1986; Bruce et al., 1987a; Ganel &

Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000), the absence of an AoA effect was not surprising. The distinctiveness effect indicated that there was something systematic in the way that male and female faces colonised MDS. Although this finding strayed from the core objective of this thesis, it opened up the possibility that MDS provides an interface between identity and gender processing, thus bolstering the subsequent exploration of the relationship between identity and gender.

Using AoA as a tool to index the nature of the relationship between identity and gender processing, Chapters 5 and 6 investigated the hypothesis that identity information affects gender processing when the visual stimulus is manipulated to remove external cues to gender. The rationale stemmed from various single-route accounts (Ganel & Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000) and parallel-route accounts (Clutterbuck & Johnston, 2004; McNeill et al., 2003; Rossion, 2002) which are united in their predictions but differ in the mechanisms purported to underpin this effect. The objective of Experiments 5 to 8 was not to elucidate mechanisms but simply to ascertain the nature of the relationship between identity and gender processing.

When AoA exerted robust effects on familiarity decisions (Experiment 5) these translated into effects on gender decisions (Experiment 6). As predicted this effect was restricted to conditions where external cues to gender had been removed. However, this result applied exclusively to male faces. These findings were not replicated in Experiments 7 and 8. Instead, performance in these familiarity decision and gender decision tasks was characterised by inconsistent findings. However, this did not constitute compelling evidence against the hypothesis that identity affects gender processing and it was concluded that mechanisms underpinning identity and gender processing are not entirely independent.

As a preliminary attempt to distinguish parallel-route and single-route explanations for the results of Experiment 6, a post-hoc analysis was carried out. The relative time course of identity and gender processing was investigated because the accounts make different predictions regarding the impact of this factor on gender decision latencies. As the single-route hypothesis predicts an effect of identity on gender processing when the perceptual whole is truncated (when external features are removed), altering the relative time course of the operations should not modulate effects of identity on gender decisions. Conversely, parallel-route accounts predict that whenever gender processing is slowed sufficiently, whether or not this is a result of androgyny or the removal of external features, other processes may facilitate gender decisions. As such, these accounts predict that altering the relative time course of the operations should modulate effects of identity on gender processing. In the whole face condition, when gender decisions were relatively slow and familiarity decisions fast, a trend approximating an AoA effect emerged, thus supporting parallel-route accounts. No such effects were recovered with internal features, consistent with single-route mechanisms which are not sanctioned by the relative time course of operations. Although inconclusive, this analysis opened up the possibility that effects of identity on gender processing are not simply governed by manipulation of the visual stimulus which seeks to remove external features.

Chapter 7 investigated the effect of gender typicality on the relationship between identity and gender processing as a means of exploiting discrepancies between predictions made by the parallel-route and single-route accounts. Whereas single-route mechanisms hypothesise that the relationship between identity and gender processing is modulated solely by the presence/absence of external cues, parallel-route mechanisms are affected by any manipulation which varies ease of gender processing. Therefore, an effect of gender typicality on the relationship between identity and gender processing would provide support for parallel-route accounts. Experiments 9 and 10 recruited AoA as an investigative tool to index the nature of the relationship between identity and gender processing. Effects of AoA on familiarity decisions were not comprehensive (Experiment 9), yet there was a tendency for AoA to affect gender decisions made to faces low in gender typicality (Experiment 10). This indicated that the effect of identity on gender processing may be triggered by intrinsic physiognomic variations in gender typicality in addition to other manipulations which vary ease of gender processing. These results indicated that parallel-route mechanisms are likely to underpin the effect of identity on gender processing.

Experiment 9 also demonstrated an effect of gender typicality on familiarity decisions, which taken together with the results of Experiment 10, indicated that the relationship between identity and gender processing may be one of mutual influence. The effect of gender on identity processing supported Baudouin & Tiberghien's (2002) finding that gender facilitates identification and it was also compatible with the notion that MDS interfaces identity and gender processing.

This thesis provided evidence that mechanisms underpinning AoA require no integration with those sub-serving distinctiveness effects. This independence implied that AoA and distinctiveness operate at serially-arranged processing stages located at PINs or FRU-PIN connections, and FRUs, respectively. By employing AoA as an investigative tool, this thesis also provided evidence that the cognitive operations of identity and gender processing are not entirely independent. A mutual relationship between identity and gender processing, together with an effect of distinctiveness on gender processing, indicated that MDS may provide one interface for identity and gender processing. Implications for theories of AoA and models of face processing are discussed in turn.

8.2 Implications for AoA: Mechanisms and Loci

The familiarity decision tasks reported in this thesis found robust effects of AoA, supporting other studies which uncovered effects in this domain (Lewis, 1999; Lewis et al., 2004; Moore & Valentine, 1998; 1999). There are several advantages associated with recovering AoA effects in the domain of face recognition, particularly in tasks which by-pass phonological processes. Studies in this domain demonstrate that mechanisms subserving AoA effects are not tied exclusively to lexical operations. Additionally, they provide a means of disentangling maturational mechanisms from those rooted in temporal order effects: lexical studies offer no such luxury because age and order of acquisition are inextricably linked in this domain. The present experiments support recent theories which are domain-independent and free from specific maturational constraints (A. Ellis & Lambon Ralph, 2000; Steyvers & Tenenbaum, 2005; Zevin & Seidenberg, 2002).

A. Ellis and Lambon Ralph's (2000) neural plasticity hypothesis has at its core the tenet that AoA effects are a fundamental property of the representation of perceptual and lexical information. The failure to tease out robust interactions between AoA and distinctiveness in the present series of experiments indicated that mechanisms underpinning this theory, or indeed any theory of AoA, need not be sanctioned by mechanisms associated with distinctiveness. Couched in terms of the neural plasticity hypothesis, loss of network plasticity as a result of cumulative and interleaved learning does not cause temporally-defined 'snap-shots' of multidimensional face-space to become entrenched in the network. Furthermore, it is not possible to measure the feasibility of AoA theories according to whether or not they support an interaction between AoA and distinctiveness. Had an interaction been discovered, it would have been possible to sort hypothesises into those which are capable of mechanistically supporting an interaction and those which are not.

Application of additive factors logic (Roberts & Sternberg, 1993; Sternberg, 1969; 1998) to this finding of additivity between AoA and distinctiveness led to the tentative conclusion that AoA and distinctiveness are processed by serially-arranged stages. Additionally, a number of further inferences can be made. If distinctiveness effects are modality-dependent, and AoA effects modality-specific, in that different aspects of identity (face, name, voice) may be acquired at different points in time, then it is unlikely that either variable would affect familiarity decisions at the PINs: the modality-independent site. Given that distinctiveness effects have been explicitly linked to patterns of activation at FRUs (e.g. Burton et al., 1990; Newell et al., 1999), it is argued that the AoA effect for face familiarity decisions is located in the connections between FRUs and PINs. This inference imposes some parameters on the loci of AoA effects which is particularly interesting given the school of thought that AoA effects are more or less ubiquitous.

The proposal of an integrated theory of mechanisms underpinning AoA effects for face familiarity decisions is somewhat hindered, however, by fundamental differences in the connectionist networks used to demonstrate AoA effects and those developed to model face recognition. Whilst Burton et al.'s (1990) IAC model of face recognition employed a localist architecture, A. Ellis and Lambon Ralph (2000) showed that AoA effects are a by-product of cumulative and interleaved learning in a distributed network. A. Ellis and Lambon Ralph proposed that early patterns exert greater impact on network structure than later patterns owing to a decline in network plasticity which occurs as learning proceeds. Moreover, attempts to reconfigure the network by later patterns are resisted by early patterns so long as they continue to be trained alongside the later patterns. It is conceivable how early experience affects later learning in distributed representations because units used to represent one item may be common to those representing other items. However, in a localist network, learning involves changing the connection strengths between pairs of units where each connection is peculiar to one item. Therefore in a localist network, where forging or updating a particular connection has no effect on other connections, it is hard to envisage how early experience would impact on later connections. There are no theoretical grounds to suggest that models of face recognition need be captured on one type of network over another; localist networks tend to be employed because they are easier to analyse than distributed networks. Given time, distributed networks may provide a unified account of face recognition which take into account effects of AoA.

Implications for AoA loci were also offered by the second set of experiments which sought to ascertain the nature of the relationship between identity and gender processing. This thesis argued that parallel-route accounts should be recruited to explain the influence of identity on gender processing. A subset of these accounts hypothesise that the cognitive system uses gender-relevant semantic information to aid gender decisions when gender processing is slowed (Bruce, 1986; McNeill et al., 2003; Rossion, 2002). This implies that for AoA effects to translate into gender decisions, the semantic system must provide a locus for AoA effects. This is enlightening given that some studies report effects of AoA on semantic decisions (Brysbaert et al., 2000; Lewis, 1999) whereas others do not (Moore, 1998; Morrison et al., 1992). However, a caveat should be inserted because this inference is speculative and relies on mechanisms predicted by one particular subset of parallel-route accounts.

8.3 Implications for Models of Face Processing: Heterarchic and Hierarchic Processes

Indexed by AoA, this thesis demonstrated that identity and gender appear to interact under certain circumstances. Whereas identity and gender were historically conceived as independent processes (Bruce & Young, 1986), studies increasingly report interactivity between the heterarchic operations. Proponents of parallel-route accounts hypothesise that whenever gender processing is laborious, identity information can facilitate gender decisions (Bruce, 1986; Clutterbuck & Johnston, 2004; McNeill et al., 2003; Rossion, 2002). Conversely, proponents of the single-route account state that integrality is the default arrangement with a false separability emerging between operations only when superficial external cues facilitate the decision (Ganel and Goshen-Gottstein, 2002; Goshen-Gottstein & Ganel, 2000). Both accounts predict that identity influences gender processing when external cues to gender are removed. Indeed, this thesis demonstrated that AoA affected gender decisions made to faces with external features removed, so long as faces were susceptible to AoA effects in a prerequisite familiarity decision task.

Mechanisms hypothesised to underpin this relationship appear to be chosen at the whim of researchers. Furthermore, attempts to disentangle parallel-route from single-route accounts are not possible when studies remove external cues to gender because this manipulation fosters similar predictions from both accounts. Conversely, an effect of gender typicality was reasoned to reflect the operation of parallel-route mechanisms because only these mechanisms are sympathetic to this manipulation. An effect of gender typicality therefore supported parallel-route hypotheses and results were consistent with Bruce's (1986) original hunch that "a parallel-routes model will allow the apparent influence of any one route upon another if it provides information relevant to the task in hand" (p. 396). In neurological terms, this finding is also compatible with Haxby et al.'s (2000) proposition that the fusiform gyrus mediates the processing of invariant facial information because information from one route can only be relevant to another route if it is invariant.

This evidence in favour of a parallel-route account is also consistent with an analysis at the level of perceptual representations extracted for identity and gender processing. As the representations mediating the processes are not identical, a parsimonious explanation of the relationship between the processing routes is one which deems functional independence, rather than integrality of processing as a single route.

Two recent parallel-route hypotheses offer different mechanisms to account for the effect of identity on gender processing. Rossion's (2002) account proposed that gender decisions could either be made on the basis of visually-derived semantic information or gender-relevant semantic information from the identification route. The alternative sources of gender information operate in parallel with the speed of the response dependent on which route yields the decision first. To develop the hypothesis beyond that formulated by Bruce (1986), Rossion hypothesised that the mechanisms must operate in cascade to enable the slower process of identification to influence gender processing. Alternatively, Clutterbuck and Johnston (2004) proposed that gender decisions are always made on the basis of activation from the visually-derived semantic route but there is cross-talk between this route and the identification route. The authors proposed that FRUs feedback activation to the structural encoding units to stabilise the representations they contain. Facilitation from the identification route is therefore passed down the visually-derived semantic route to speed up gender decisions.

It is argued that Rossion's (2002) account provides the most parsimonious explanation for the effect of AoA on gender decisions reported in this thesis. Clutterbuck and Johnston's (2004) hypothesis deems that activation is passed from the FRUs to the level of structural encoding. However, feedback at this level in the system is incompatible with Burton et al.'s (1990) IAC architecture which supposes that these connections are not bi-directional. Moreover, if AoA affects the connections between the FRUs and PINs, as previously hypothesised, then it is hard to see how AoA could affect gender decisions on the basis of FRU feedback alone. Clutterbuck and Johnston's account does not provide a parsimonious explanation.

This thesis also provided evidence that the relationship between identity and gender processing is one of mutual influence with gender typicality affecting familiarity decisions. Although Clutterbuck and Johnston (2004) reasoned that it is hard to imagine how gender information could ever assist identification, Rossion (2002) suggested that a cascade system could allow a two-way interaction between the two cognitive operations. However, he failed to specify how this effect might be achieved. Baudouin and Tiberghien (2002) proposed a hypothesis to explain how information from the visually-derived semantic route could offer information for relevant to identification. Made possible by studies which suggest that MDS is systematically colonised into male and female regions (Experiment 4 in this thesis; Johnston et al., 1997; O'Toole et al., 1998), Baudouin and Tiberghien argued that visually-derived gender information enables the search for a particular identity in MDS to be limited to an area of space corresponding to the perceived gender. The authors hypothesised that the cognitive system is responsible for the subdivision of MDS. As an extrapolation of this hypothesis, this thesis argues that the more quickly visually-derived gender information is made available to the cognitive system, the faster this information can be directed to MDS to facilitate identification. Hence why familiarity decisions were faster to gender typical faces compared with faces low in gender typicality.

Implications for the hierarchic processes involved in face recognition stemmed from the finding that AoA and distinctiveness are independent. Whereas Bruce and Young (1986) proposed that familiarity decisions are taken at the level of FRUs, subsequent models suggest that PINs accommodate this decision (Burton et al., 1990; Burton et al., 1999). The tentative inference that distinctiveness affects activity at FRUs and AoA modulates the connections between FRUs and PINs, is most parsimoniously accommodated by a PIN-level familiarity decision mechanism. Additionally, neuroimaging evidence supports the contention that familiarity decisions are taken at the level of modality-independent PINs (Shah et al., 2001).

It might seem tautological to resurrect resolved issues, because the PIN-level familiarity decision mechanism is widely accepted, but Goshen-Gottstein and

Ganel (2000) still maintain that familiarity decisions are taken at the level of FRUs: to permit repetition priming effects in tasks other than those requiring identity information, the authors reallocated familiarity decisions to FRUs and hypothesised that FRUs support gender decisions and can be formed after a single presentation of a face. In addition to supporting a PIN-level familiarity decision mechanism, this thesis supplies further evidence to wage against the single-route hypothesis.

8.4 Conclusions

Together with studies in the domains of word and object recognition, broadly speaking this thesis confirms our understanding of AoA as a variable underpinned by domain-independent mechanisms which are not sanctioned by maturational constraints. Just as models in the other domains should take into account AoA effects, models of face processing should seek to simulate the effect. After all, the ability of face processing models to accommodate a continually growing range of psychological phenomena will inform an increasingly detailed understanding of mechanisms underpinning face processing. In this thesis, robust effects of AoA on familiarity decisions and various effects on gender decisions offered implications for operations involved in face processing in addition to the variable itself.

That AoA did not interact with distinctiveness in familiarity decision tasks indicated that theories of AoA and distinctiveness require no elaboration. Moreover, the additivity of the two variables implied, albeit speculatively, that different stages in processing accommodate AoA and distinctiveness, thus imposing constraints on the loci of AoA effects. The inference that AoA affects the connections between the FRUs and the PINs and distinctiveness effects operate at the level of FRUs also supports models of face processing which attribute familiarity decisions to PINs.

That identity influenced gender processing when gender decisions could not be made on the basis of superficial heuristics, indicated that the relationship between the cognitive operations is not entirely independent. Moreover, parallel-route accounts were demonstrated to be the most likely candidate to explain this relationship. An unanticipated and somewhat inconsistent effect of gender on identification demonstrated that the relationship between identity and gender processing may be best described as reciprocal. Although it is hard to conceive how gender information may aid identification, this finding is compatible with the notion that MDS interfaces identity and gender processing.

8.5 Future Work

8.5.1 Effects of Age-Related Changes on AoA

Little is known about whether or not face representations are ageless and to what extent recognition demands an exact match of age characteristics. Also, it is not known whether or not AoA effects accrue as faces undergo age-related changes which may or may not cause deviation from their original representations. In a recognition memory task, George and Hole (1998) demonstrated that unfamiliar faces were recognised despite age transformations. However, performance was impaired indicating that representations are not ageless. The authors concluded that "age information is utilised in a broad way within a representation, perhaps from some kind of prototypical age face" (p. 1132). Compatible with this suggestion, studies have demonstrated that faces become increasingly distinctive with age (Deffenbacher et al., 1998; O'Toole et al, 1997). This implies that representations shift in MDS over time. Studies of AoA in the domain of face recognition are incomplete without considering the impact of age-transformations, given the temporal dimension intrinsic to AoA.

In the present experiments, and indeed previous face recognition studies, AoA is confounded with degree of age-transformation. Regardless of whether or not up-to-date images or dated images are used, early acquired faces have changed since acquisition to a greater degree than late acquired faces. It is possible that an up-to-date image of an early acquired face will not tap the representation of the face when it was acquired and a dated image of an early acquired face will hamper recognition of that face. In both cases the emergence of an AoA effect could be compromised. To remove this potential confound pragmatically, future studies could employ one set of faces and vary age of participant. In this way celebrities who have risen to fame recently will be early acquired to a younger group of participants and late acquired to an older group, and age-transformation will not be confounded with AoA. However, to banish this confound on theoretical grounds rather than tackle the issue pragmatically would reflect a step forward in our understanding of the impact of age-related changes on AoA.

8.5.2 Mechanisms Underpinning Parallel-Route Accounts

This thesis concluded that parallel-route accounts, in particular the account offered by Rossion (2002), provided the most likely explanation for the effect of identity on gender processing found in Experiments 6 and 10. Rossion developed Bruce's (1986) original notion that gender-relevant semantic information could feed gender decisions within the cognitive system sufficiently quickly to show an apparent effect of identity on gender processing. However, Rossion proposed that these processes need to operate in cascade so that "later processes may be initiated before earlier ones are completed and distant stages may inhibit or facilitate adjacent or more distant ones" (p. 1015). Clutterbuck and Johnston (2004) hypothesised that the effect of identity on gender processing was not the product of a decision-making process at the cognitive system; rather it was the direct result of communication between the processing routes themselves. The authors proposed that FRUs feedback activation to the structural encoding units to stabilise the representations they contain. Therefore, unfamiliar faces rely solely on the analysis of facial structure whereas familiar faces receive structural information facilitated by FRU feedback.

Increasingly, studies are reporting effects of identity on gender processing leading to the conclusion that the heterarchic operations are not entirely independent. To inform models of face processing, an understanding of the mechanisms mediating this effect is vital. Although this thesis considers Rossion's (2002) account to be more parsimonious than Clutterbuck and Johnston's (2004) hypothesis, particularly in IAC terms, research directly aimed at disentangling the accounts is warranted. More specifically, this research should seek to elucidate whether or not interactivity stems from decision-making processes at the cognitive system or direct communication between the two processing routes themselves.

8.5.3 Problems Inherent in Face Recognition Studies

Several problems with studies that employ famous faces as their stimuli were highlighted by Lewis (1999a) and apply to the present experiments. Firstly, there is the problem of obtaining enough stimuli sufficiently familiar to the majority of participants. This is compounded by the idiosyncratic nature of celebrity knowledge: despite being rated as familiar, Ewan McGregor was only recognised by 43% of participants in Experiment 2. However, in Experiment 9 86% of participants were able to correctly identify his face. Secondly, Lewis noted the difficulties involved in standardising the images so that they are equally good depictions of the faces. Although 'likeness' was matched across sets in the present experiments, this does not constitute standardisation. It simply ensures that there are equal numbers of poor representations and good representations across the sets. Further, the transient nature of faces causes several problems. Faces may fall out of the public eye and popularity may wane unpredictably, and particular faces may be susceptible to cohort effects. Consequently, variables which affect recognition, such as frequency and familiarity, may vary as a function of media trends or the cohort.

These problems standardising recognisability may have been responsible for the disappointing lack of robust effects in this thesis. Unfortunately, the more time that elapses since the creation of the databases, the more prone to artefactual effects future experiments will become. To resolve inconsistencies, namely the failure to replicate Experiment 6, and to combat weak findings, namely the lack of significant AoA effects in Experiment 10, it would be interesting to pursue this line of research with new databases created and rated to reflect current media trends. Unlike words and pictures, faces are dynamic. Unfortunately, this means that famous face recognition studies need to dedicate copious amounts of time to collecting stimuli and keeping them up-to-date.

REFERENCES

Alley, T.R. (1988). *Social and applied aspects of perceiving faces.* Lawrence Erlbaum Associates.

Barry, C., Morrison, C.M., & Ellis, A.W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology, 50A,* 560-585.

Bartlett, J.C., Hurrey, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12,* 219-228.

Baudouin, J.Y., & Tiberghien, G. (2002). Gender is a dimension of face recognition. *Journal of Experimental Psychology, 28,* 362-365.

Bonner, L., & Burton, A.M. (2003). Getting to know you: How we learn new faces. *Visual Cognition, 10,* 527-536.

Brown, E., & Perrett, D.I. (1993). What gives a face its gender? *Perception, 22,* 829-840.

Brown, G.D.A., & Watson, F.L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15,* 208-216.

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

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception, 15,* 387-397.

Bruce, V. (1983). Recognising Faces. *Philosophical Transactions of the Royal Society, 302B,* 423-436.

Bruce, V., Burton, A.M., & Dench, N. (1994). What's distinctive about a distinctive face? *Quarterly Journal of Experimental Psychology, 47A,* 119-141.

Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., et al. (1993a). Sex discrimination: how do we tell the difference between male and female faces? *Perception, 22,* 131-152.

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

Bruce, V., Ellis, H., Gibling, F., & Young, A. (1987a). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology, 41,* 510-520.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23,* 803-822.

Bruce, V., Valentine, T., & Baddeley, A. (1987b). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology, 1,* 109-120.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104,* 215-226.

Burton, A.M., & Vokey, J.R. (1998). The face-space typicality paradox: Understanding the face space metaphor. *The Quarterly Journal of Experimental Psychology, 51A,* 476-483.

Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22,* 153-176.

Burton, A.M., Bruce, V., & Hancock, P.J.B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science, 23,* 1-31.

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81,* 316-380.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C.R., McGuire, P.K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276,* 593-596.

Carroll, J.B., & White, M.N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25,* 85-95.

Chang, P.P.W., Levine, S.C., & Benson, P.J. (2002). Children's recognition of caricatures. *Developmental Psychology, 38,* 1038-1051.

Clutterbuck, R., & Johnston, R.A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 31,* 985-994.

Clutterbuck, R., & Johnston, R.A. (2004). Demonstrating the acquired familiarity of faces by using a gender-decision task. *Perception, 33,* 159-168.

Cohen, J. (1988). *Statistical power analyses for the behavioural sciences.* Hillsdale, NJ: Erlbaum.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behaviour Research Methods, Instruments, and Computers, 25,* 257-271.

Craw, I. (1995). A manifold model of face and object recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition: Explorations in face space.* London: Routledge.

Deffenbacher, K.A., Vetter, T., Johanson, J., & O'Toole, A.J. (1998). Facial aging, attractiveness, and distinctiveness. *Perception, 27,* 1233-1243.

Dubois, S., Rossion, B., Schiltz, C., Bodart, J.M., Michel, C., Bruyer, R., et al. (1999). Effect of familiarity on the processing of human faces. *NeuroImage, 9,* 278-289.

Ellis, H.D. (1992). The development of face processing skills. *Philosophical Transactions of the Royal Society of London, 335B,* 105-111.

Ellis, A.W., & Lambon Ralph, M. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1103-1123.

Ellis, A.W., Burton, A.M., Young, A.W., & Flude, B.M. (1997). Repetition priming between parts and wholes: Tests of a computational model of familiar face recognition. *British Journal of Psychology, 88,* 579-608.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of faces recognition. *Perception, 8,* 431-439.

Ellis, H.D., Shepherd, J.W., Gibling, F., & Shepherd, J. (1988). Stimulus factors in learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Ed.), *Practical Aspects of Memory, Current Research and Issues: Vol. 1. Memory in Everyday Life* (pp. 136-143). Chichester, UK: Wiley.

Enlow, D. (1982). *Handbook of facial growth.* Philadelphia: W.H. Saunders.

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrity of sex and identity of faces: Further evidence for the single-route hypothesis. *Journal of Experimental Psychology, 28,* 854-867.

George, P.A., & Hole, G.J. (1998). Recognising the ageing face: The role of age in face processing. *Perception, 27,* 1123-1134.

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 267-283.

Ghyselinck, M., Lewis, M.B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica, 115,* 43-67.

Gilhooly, K.J., & Gilhooly, M.L.M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology, 71,* 105-110.

Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgement task: Evidence for a common route for the processing of sex and identity. *Journal of Experimental Psychology, 26,* 1198-1214.

Hancock, P.J.B., Burton, A.M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & Cognition, 24,* 26-40.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223-233.

Hay, D.C., & Young, A.W. (1982). The human face. In A.W. Ellis (Ed.), *Normality and pathology in cognitive functions.* London: Academic Press.

Hosie, J.A., & Milne, A.B. (1995). Distinctiveness and memory for unfamiliar faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition: Explorations in face space.* London: Routledge.

Johnston, R.A., & Ellis, H.D. (1995). Age effects in the processing of typical and distinctive faces. *The Quarterly Journal of Experimental Psychology, 48A,* 447-465.

Johnston, R.A., Kanazawa, M., Kato, T., & Oda, M. (1997). Exploring the structure of multidimensional face-space: The effects of age and gender. *Visual Cognition, 4,* 39-57.

Kirby, M., & Sirovich, L. (1990). Applications of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12,* 103-108.

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology, 99,* 199-208.

Langlois, J.H., & Roggman, L.A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115-121.

Leveroni, C.L., Seidenberg, M., Mayer, A.R., Mead, L.A., Binder, J.R., & Rao, S.M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience, 20,* 878-886.

Lewis, M.B., & Johnston, R.A. (1999). A unified account of the effects of caricaturing faces. *Visual Cognition, 6,* 1-41.

Lewis, M.B. (1999a). Age of acquisition in face categorisation: is there an instance-based account? *Cognition, 71,* B23-B29.

Lewis, M.B. (1999b). Are age-of-acquisition effects cumulative-frequency effects in disguise? A reply to Moore, Valentine and Turner (1999). *Cognition, 72,* 311-316.

Lewis, M.B. (2004). Face-space-R: Towards a unified account of face recognition, *Visual Cognition, 11,* 29-69.

Lewis, M.B., Chadwick, A.J., & Ellis, H.D. (2002). Exploring a neural-network account of age-of-acquisition effects using repetition priming of faces. *Memory & Cognition, 30,* 1228-1237.

Lewis, M.B., Gerhand, S., & Ellis, H.D. (2001). Re-evaluating age-of acquisition effects: are they simply cumulative-frequency effects? *Cognition, 78,* 189-205.

Light, L.L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 212-228.

McNeill, A., Burton, A.M., & Ellis, A.W. (2003). When sex isn't special: Priming onto a sex decision. *Visual Cognition, 10,* 641-650.

Monaghan, J., & Ellis, A.W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 183-206.

Moore, V., & Valentine, T. (1998). The effect of age of acquisition on speed and accuracy of naming famous faces. *The Quarterly Journal of Experimental Psychology, 51A,* 485-513.

Moore, V., & Valentine, T. (1999). The effects of age of acquisition in processing famous faces and names: exploring the locus and proposing a mechanism. In *Proceedings of the twenty-first annual meeting of the cognitive science society, 416-421.* New Jersey: Erlbaum.

Moore, V., Valentine, T., & Turner, J. (1999). Age-of-acquisition and cumulative frequency have independent effects. *Cognition, 72,* 305-309.

Morrison, C.M., Ellis, A.W., & Quinlan, P.T. (1992). Age of acquisition, not word frequency, affects object naming not object recognition. *Memory & Cognition, 20,* 705-714.

Morrison, C.M., Chappell, T.D., & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50A,* 528-559.

Moscovitch, M., Goshen-Gottstein, Y., & Vriezen, E.R. (1993). Memory without conscious recollection: A tutorial review from a neuropsychological perspective. In C. Umilta & M. Moscovitch (Ed.), *Attention and Performance XV: Conscious and nonconscious processes in cognition.* Cambridge, MA: MIT Press.

Newell, N.N., Chiroro, P., & Valentine, T. (1999). Recognising unfamiliar faces: The effects of distinctiveness and view. *The Quarterly Journal of Experimental Psychology, 52A,* 509-534.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception, 30,* 755-764.

O'Toole, A.J., Deffenbacher, K.A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition, 26,* 146-160.

O'Toole, A.J., Vetter, T., Volz, H., & Salter, E.M. (1997). Three dimensional caricatures of human heads: Distinctiveness and the perception of facial age. *Perception, 26,* 719-732.

Oldfield, R.C., & Wingfield, A. (1964). The time it takes to name and object. *Nature, 202,* 1031-1032.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19,* 473-497.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D.E. Meyer & S. Kornblum (Ed.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence and Cognitive Neuroscience.* Cambridge, MA: MIT Press.

Roberts, T., & Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces. *Perception, 17,* 475-481.

Rossion, B. (2002). Is sex categorization from faces really parallel to face recognition? *Visual Cognition, 9,* 1003-1020.

Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces? A PET study of face categorical perception. *Journal of Cognitive Neuroscience, 13,* 1019-1034.

Sadr, J., Jarundi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception, 32,* 285-293.

Sergent, J., Otha, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain, 115,* 15-36.

Shah, N., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., et al. (2001). The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain, 124,* 804-815.

Smith, M.A., Cottrell, G.W., & Anderson, K.L. (2001). The early word catches the weights. In T.K. Leen, T.G. Dietterich, & V. Tresp (Ed.), *Advances in neural information processing systems.* Cambridge, MA: MIT Press.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donders' method. In W.G. Koster (Ed.), *Attention and performance II.* Amsterdam: North Holland Press.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Ed.), *An invitation to cognitive science: methods, models, and conceptual issues.* Cambridge, MA: MIT Press.

Steyvers, M., & Tenenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41-78.

Taft, M., & van Graan, F. (1998). Lack of phonological mediation in a semantic categorization task. *Journal of Memory and Language, 38,* 203-224.

Tranel, D., Damasio, A.R., & Damasio, H. (1988). Intact recognition of facial expression, sex, and age in patients with impaired recognition of face identity. *Neurology, 28,* 690-696.

Tuck, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71-86.

Turner, J.E., Valentine, T., & Ellis, A.W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition, 26,* 1282-1291.

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognizing and classifying faces. *Perception, 15,* 525-535.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology, 44A,* 671-703.

Valentine, T. (1999). Face-space models of face recognition. In Wenger, M.J. & Townsend, J.T. (Ed.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges.* Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Valentine, T., Hollis, J., & Moore, V. (1998). On the relationship between reading, listening and speaking: It's different for people's names. *Memory and Cognition, 26,* 740-753.

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

Vokey, J.R., & Read, J.D. (1995). Memorability, familiarity, and categorical structure in the recognition of faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition: Explorations in face space.* London: Routledge.

Wickham, L.H.V., Morris, P.E., & Fritz, C.O. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology, 91,* 99-123.

Wild, H.A., Barrett, S.E., Spence, M.J., O'Toole, A.J., Cheng, Y.D., & Brooke, J. (2000). Recognition and sex categorisation of adults' and childrens' faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology, 77,* 269-291.

Yamazaki, M., Ellis, A.W., Morrison, C.M., & Lambon Ralph, M.A. (1997). Two acquisition effects in the reading of Japanese Kanji. *British Journal of Psychology, 88,* 407-421.

Young, A.W., Hay, D.C., & Ellis, A.W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology, 76,* 495-523.

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

Zevin, J.D., & Seidenberg, M.S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47,* 1-29.

APPENDICES

Appendix 1. *Stimuli from the Pilot Experiment.*

Condition	Stimulus	Ratings				
		AoA	Distinctiveness	Frequency	Average no. facts	Likeness
Early acquired distinctive	Barbara Windsor	2.80	4.20	3.80	6.75	4.10
	Elton John	2.25	3.45	3.05	7.85	4.30
	Goldie Hawn	2.61	3.35	1.95	5.42	3.84
	George Michael	2.80	3.75	3.35	7.45	4.10
	Julia Roberts	2.83	3.80	2.45	5.55	4.05
	Jimmy Saville	2.12	4.35	1.75	6.00	3.00
	John Travolta	2.25	3.30	2.50	6.35	4.68
	Madonna	2.33	3.60	3.00	8.40	4.50
	Michael Jackson	1.76	4.90	2.20	8.35	3.90
	Oprah Winfrey	2.83	4.00	2.65	5.50	3.63
	Prince Charles	1.60	3.80	3.75	8.35	4.70
	Prince	2.88	4.70	1.88	6.10	3.50
	Queen Mother	1.50	3.60	3.10	5.95	4.30
	Rowan Atkinson	2.52	4.75	2.75	6.80	4.65
	Rod Stewart	2.83	4.10	1.80	4.60	3.74
	Sean Connery	2.68	3.80	2.60	6.40	3.80
Late acquired distinctive	Bill Clinton	3.60	3.70	3.30	7.10	4.90
	Billy Connelly	3.11	4.25	2.30	5.70	4.15
	Chris Evans	3.36	3.70	3.60	7.70	4.00
	David Bowie	2.95	3.90	2.40	5.80	4.00
	Geri Halliwell	3.72	3.55	3.05	7.75	4.35
	Jamie Oliver	3.48	4.25	2.67	7.35	3.95
	Janet Jackson	3.05	3.90	2.07	5.35	3.80
	Jim Carey	4.18	3.25	3.35	6.90	3.80
	Mel Gibson	2.94	3.35	2.65	5.20	4.40
	Naomi Cambell	3.24	3.85	2.42	4.85	3.60
	Pamela Anderson	2.95	3.55	2.65	7.60	4.30
	Pierce Brosnan	3.71	3.35	2.60	5.20	4.75
	Richard Gere	3.04	3.70	2.05	5.95	3.35
	Robbie Williams	3.16	3.65	3.74	8.25	4.05
	Sylvester Stallone	2.94	4.30	2.10	5.45	3.75
	Tony Blair	3.84	3.30	4.42	8.25	4.05
Early acquired typical	Bob Monkhouse	2.42	3.16	3.15	5.10	4.65
	Cliff Richard	2.57	3.05	2.25	6.95	2.74
	Carol Vorderman	2.63	2.75	4.20	6.84	3.30
	Des Lynam	2.60	3.20	3.47	5.85	4.75
	Esther Rantzen	2.44	3.15	2.40	5.50	4.15
	Harrison Ford	2.45	3.00	1.94	5.85	4.00
	John Cleese	2.50	3.20	2.80	6.75	4.60
	Judy Finnigan	2.90	2.50	3.35	6.70	3.75
	John Major	2.75	2.80	2.55	7.00	4.70
	Kylie Minogue	2.00	2.50	2.95	7.85	2.63

Appendix 1. *Stimuli from the Pilot Experiment – Continued.*

Condition	Stimulus	Ratings				
		AoA	Distinctiveness	Frequency	Average no. facts	Likeness
Early acquired typical	Paul McCartney	2.71	2.85	2.65	9.00	3.65
	Queen	1.32	3.25	3.25	6.35	4.85
	Sarah Ferguson	2.44	2.45	2.40	6.45	4.20
	Tom Hanks	2.88	3.25	2.47	7.70	4.10
	Terry Wogan	2.04	2.85	2.63	6.26	4.85
	Victoria Wood	2.61	2.25	2.35	5.80	3.40
Late acquired typical	Britney Spears	4.52	2.15	3.07	6.90	4.35
	Celine Dion	3.26	3.10	2.35	6.25	3.45
	David Beckham	3.80	2.74	3.75	9.15	4.55
	Gwyneth Paltrow	3.82	2.90	2.47	8.05	4.45
	Hilary Clinton	3.94	2.85	2.44	6.65	3.50
	Jerry Springer	4.08	2.90	3.58	5.85	4.00
	Kate Winslett	3.95	2.55	2.41	6.40	3.35
	Meg Ryan	3.06	2.85	2.07	4.95	3.10
	Michelle Pfeiffer	3.30	2.65	2.47	5.25	3.80
	Natalie Imbruglia	3.40	3.25	2.65	7.10	4.35
	Nicole Kidman	3.26	2.85	2.37	5.85	3.40
	Ricky Martin	3.20	2.10	3.65	7.25	4.60
	Ronan Keating	4.07	2.95	3.15	5.60	4.50
	Sharon Stone	3.56	2.60	2.29	4.85	3.32
	Sting	3.16	2.65	1.95	6.95	4.10
	Victoria Beckham	3.92	3.20	3.60	9.20	4.30

Appendix 2. *Stimuli from Experiment 1.*

Condition	Stimulus	Ratings			
		AoA	Distinctiveness	Frequency	Likeness
Early acquired distinctive	Arnold Schwarzenegger	2.67	3.90	2.05	3.81
	Barbara Streisand	3.24	3.25	1.70	4.00
	Billy Connelly	3.29	4.45	2.45	4.15
	Bob Monkhouse	2.81	3.30	2.45	4.19
	Celine Dion	3.38	2.80	2.00	3.24
	Dawn French	3.14	3.10	2.60	3.62
	Des Lynam	2.90	2.80	2.40	3.90
	Gary Lineaker	2.81	3.20	2.75	4.10
	George Michael	3.14	3.15	2.15	3.43
	Goldie Hawn	3.24	3.40	2.15	3.40
	Janet Jackson	3.29	3.15	2.15	4.24
	Joan Collins	3.38	4.10	2.00	3.95
	John Travolta	2.95	2.90	2.45	4.14
	Kylie Minogue	2.33	3.30	3.95	4.29
	Lenny Henry	3.24	2.85	2.35	3.67
	Leslie Joseph	3.09	3.70	2.05	3.95
	Madonna	2.62	3.25	3.80	3.67
	Margaret Thatcher	2.33	3.05	1.95	3.62
	Michael Jackson	2.14	4.80	3.65	4.62
	Oprah Winfrey	3.38	2.90	2.10	3.90
	Pamela Anderson	3.19	3.60	2.30	3.76
	Philip Schofield	2.52	2.83	3.39	4.00
	Prince Charles	2.14	3.65	3.15	3.95
	Prince Phillip	2.71	2.85	2.25	4.00
	Princess Diana	2.26	3.04	2.61	3.74
	Queen	1.48	3.25	3.65	3.43
	Richard Wilson	3.00	3.13	2.35	3.91
	Robin Williams	2.90	3.05	2.50	3.67
	Rowan Atkinson	2.38	3.90	2.20	3.67
	Sylvester Stallone	3.10	3.95	1.68	3.05
	Terry Wogan	2.76	2.95	2.25	3.95
	Tom Hanks	2.86	3.20	2.70	3.67
	Tom Jones	3.00	3.30	2.35	4.19
	Whitney Houston	3.14	3.05	2.25	3.76
Late acquired distinctive	Alan Titchmarsh	4.18	3.26	3.04	3.77
	Calista Flockhart	4.14	3.25	3.00	4.33
	Caroline Quentin	3.90	3.35	2.41	3.05
	Chris Evans	3.57	3.65	1.90	3.81
	Dale Winton	3.67	3.35	2.50	3.81
	Dustin Hoffman	3.83	3.22	1.71	3.45
	Edwina Currie	4.32	3.35	1.53	3.90
	Fiona Bruce	4.45	3.26	2.94	3.58

Appendix 2. *Stimuli from Experiment 1 – Continued.*

Condition	Stimulus	Ratings			
		AoA	Distinctiveness	Frequency	Likeness
Late acquired distinctive	Halle Berry	4.71	3.55	3.05	4.52
	Hugh Laurie	3.67	4.13	1.95	3.61
	Jeremy Clarkson	3.81	3.57	2.64	3.86
	Jeremy Paxman	4.10	3.57	2.70	4.05
	Johnny Ball	3.57	3.13	1.67	3.31
	Lawrence Llewellyn Bowen	4.29	3.00	2.90	3.76
	Leslie Ash	4.00	3.40	2.75	3.35
	Lisa Riley	3.95	3.45	2.80	4.05
	Liv Tyler	4.35	3.15	2.70	4.24
	Liz Hurley	3.76	3.00	3.15	4.33
	Martin Clunes	3.62	3.90	2.35	4.10
	Michael Palin	3.67	2.96	2.14	3.95
	Naomi Campbell	3.52	3.45	2.45	4.38
	Natalie Imbruglia	3.67	3.15	2.55	4.62
	Peter Stringfellow	4.10	3.85	1.60	3.81
	Rene Zellweger	4.57	3.05	2.90	3.43
	Ricky Martin	4.19	3.20	1.85	3.81
	Russell Crowe	4.67	2.85	2.75	4.05
	Sandra Bullock	3.76	2.95	2.75	4.19
	Sophie Ellis Bextor	4.71	4.35	2.65	4.00
	Stephen Fry	3.81	3.91	2.73	3.38
	Sue Nicholls	3.56	2.90	2.35	3.94
	Tony Blackburn	4.37	3.91	2.26	3.21
	Vinnie Jones	3.90	3.40	2.47	4.20
	Will Smith	3.65	3.70	3.10	4.62
	Will Young	5.00	3.20	3.10	3.95
Early acquired typical	Alan Shearer	3.22	2.43	2.26	3.64
	Ant McPartlin	2.38	2.35	1.85	4.29
	Anthea Turner	2.81	2.10	3.10	4.24
	Carol Vorderman	3.10	2.60	3.05	3.33
	Danni Minogue	2.95	2.60	3.40	3.81
	Declan Donnelly	2.81	1.95	3.10	4.29
	Delia Smith	3.38	1.55	2.20	3.67
	Derek Thompson	3.20	1.39	2.24	3.50
	Diane Louise Jordan	2.65	2.70	1.85	3.94
	Esther Rantzen	3.00	2.65	2.00	3.71
	Gaby Roslin	3.00	2.65	2.61	3.87
	Harrison Ford	2.71	2.45	2.40	4.05
	Jennifer Saunders	3.38	2.00	2.20	3.43
	Jim Davidson	2.96	2.57	2.22	4.04
	John Leslie	2.38	1.65	2.50	4.29
	Keith Chegwin	3.35	2.39	2.14	3.89
	Kevin Costner	3.29	2.70	2.00	3.52
	Les Dennis	3.29	1.70	2.20	3.86
	Matthew Kelly	3.00	2.52	2.74	3.77
	Mel Gibson	3.29	2.65	2.40	4.24
	Michael Barrymore	2.81	1.90	2.42	4.05
	Michael Buerke	3.38	2.65	3.14	4.10

Appendix 2. *Stimuli from Experiment 1 – Continued.*

Condition	Stimulus	Ratings			
		AoA	Distinctiveness	Frequency	Likeness
Early acquired typical	Patricia Routledge	3.06	2.70	2.19	4.24
	Paul McCartney	2.38	2.45	2.00	4.10
	Pauline Quirke	2.33	2.50	2.75	4.05
	Philippa Forrester	3.27	2.22	2.14	3.95
	Prince William	3.19	2.45	1.90	3.40
	Princess Anne	2.76	2.45	3.25	3.57
	Richard Gere	3.33	2.30	2.25	2.48
	Ronan Keating	3.41	1.83	2.57	3.00
	Sarah Ferguson	2.76	2.00	1.90	3.38
	Sting	2.81	2.70	2.35	4.00
	Todd Carty	3.25	2.52	2.91	4.21
	Wendy Richard	2.67	2.15	3.45	4.24
Late acquired typical	Anna Kournikova	4.29	2.25	2.50	2.90
	Anne Robinson	4.00	2.45	3.80	4.52
	Bob Mortimer	3.68	2.13	2.57	3.70
	Carol Smilie	3.71	2.45	2.95	4.52
	Cat Dealy	4.43	2.55	3.80	4.24
	Darius	4.90	2.30	3.15	4.19
	David Duchovny	3.50	2.10	2.10	3.62
	David Seaman	3.52	2.45	2.05	4.24
	Denise Van Outen	3.81	2.45	3.25	4.05
	Donna Air	4.00	1.60	2.06	2.89
	Gillian Anderson	3.52	2.40	2.05	3.57
	Holly Valance	4.76	2.50	3.20	4.14
	Jill Dando	3.83	2.13	1.43	4.00
	Johnny Vaughan	3.95	1.80	2.65	4.00
	Kate Winslett	3.95	2.30	2.75	4.00
	Kenneth Branagh	3.94	2.30	1.59	2.95
	Leonardo Di Caprio	3.81	2.55	3.00	4.38
	Mathew Perry	3.81	2.55	3.45	4.19
	Michael Owen	4.24	1.60	3.15	4.14
	Michelle Collins	3.67	2.15	2.10	4.00
	Neil Kinnock	3.58	1.75	1.38	3.89
	Neil Morrissey	3.57	2.00	2.10	4.00
	Nigel Havers	4.25	2.00	1.56	4.00
	Rachel Stevens	4.43	2.25	3.25	4.14
	Ricky Gervais	4.95	2.35	2.85	3.90
	Rona Cameron	4.71	2.25	2.35	3.89
	Ross Kemp	3.62	2.25	2.00	4.71
	Sharon Stone	3.60	2.40	1.90	2.75
	Simon Cowell	4.95	2.30	2.80	3.86
	Steve Coogan	4.20	2.25	1.84	3.00
	Sue Barker	3.67	2.13	2.59	3.50
	Sue Perkins	4.60	2.25	2.65	3.75
	Terry Venables	3.76	2.00	2.00	3.67
	Vic Reeves	3.95	2.50	2.20	2.81

Appendix 3. *Stimuli from Experiments 2 and 4.*

		Ratings			
Condition	Stimulus	AoA	Distinctiveness	Frequency	Likeness
Early acquired distinctive	Arnold Schwarzenegger	2.67	3.90	2.05	3.81
	Cilla Black	2.38	3.50	2.55	4.62
	David Bowie	2.76	3.70	1.80	3.43
	David Jason	2.71	3.40	2.75	4.67
	Elton John	2.67	4.05	3.10	4.00
	Gary Lineaker	2.81	3.20	2.75	4.10
	Goldie Hawn	3.24	3.40	2.15	3.40
	John Travolta	2.95	2.90	2.45	4.14
	Kylie Minogue	2.33	3.30	3.95	4.29
	Leslie Joseph	3.09	3.70	2.05	3.95
	Margaret Thatcher	2.33	3.05	1.95	3.62
	Michael Jackson	2.14	4.80	3.65	4.62
	Pat Sharp	2.73	3.04	1.18	3.15
	Princess Diana	2.26	3.04	2.61	3.74
	Rowan Atkinson	2.38	3.90	2.20	3.67
	Sean Connery	2.71	3.30	2.40	4.43
	Sylvester Stallone	3.10	3.95	1.68	3.05
	Terry Wogan	2.76	2.95	2.25	3.95
Late acquired distinctive	Billy Piper	4.00	2.75	1.70	3.90
	Chris Evans	3.57	3.65	1.90	3.81
	Hugh Laurie	3.67	4.13	1.95	3.61
	Jeremy Clarkson	3.81	3.57	2.64	3.86
	Judi Dench	3.48	3.09	2.65	4.48
	Lawrence Llewellyn Bowen	4.29	3.00	2.90	3.76
	Lisa Riley	3.95	3.45	2.80	4.05
	Liv Tyler	4.35	3.15	2.70	4.24
	Martin Clunes	3.62	3.90	2.35	4.10
	Meg Ryan	3.48	2.80	2.25	3.25
	Michael Caine	3.48	2.75	2.25	4.10
	Naomi Campbell	3.52	3.45	2.45	4.38
	Pierce Brosnan	3.95	2.75	2.55	4.24
	Ricky Martin	4.19	3.20	1.85	3.81
	Sophie Ellis Bextor	4.71	4.35	2.65	4.00
	Stephen Fry	3.81	3.91	2.73	3.38
	Vinnie Jones	3.90	3.40	2.47	4.20
	Will Young	5.00	3.20	3.10	3.95
Early acquired typical	Alan Shearer	3.22	2.43	2.26	3.64
	Anthea Turner	2.38	2.35	1.85	4.29
	Carol Vorderman	3.10	2.60	3.05	3.33
	Danni Minogue	2.95	2.60	3.40	3.81
	Gary Barlow	2.90	1.74	1.32	3.33
	Harrison Ford	2.71	2.45	2.40	4.05

Appendix 3. *Stimuli from Experiments 2 and 4 – Continued.*

Condition	Stimulus	Ratings			
		AoA	Distinctiveness	Frequency	Likeness
Early acquired typical	John Leslie	2.38	1.65	2.50	4.29
	Keith Chegwin	3.35	2.39	2.14	3.89
	Les Dennis	3.29	1.70	2.20	3.86
	Linda Robson	2.33	1.80	1.70	3.76
	Michaela Strachen	2.00	2.43	1.61	4.05
	Michael Barrymore	2.81	1.90	2.42	4.05
	Pauline Quirke	2.38	2.45	2.00	4.10
	Ronan Keating	3.41	1.83	2.57	3.00
	Todd Carty	3.25	2.52	2.91	4.21
	Victoria Wood	2.95	2.65	1.58	4.33
	Wendy Richard	2.67	2.15	3.45	4.24
	Zoe Ball	3.43	2.55	2.95	4.19
Late acquired typical	Angus Deayton	3.76	2.65	2.40	3.62
	Anna Kournikova	4.29	2.25	2.50	2.90
	Cameron Diaz	3.95	2.65	2.90	4.00
	Ewan McGregor	3.95	2.25	2.75	3.52
	Harry Enfield	3.62	2.55	2.45	3.38
	Jill Dando	3.83	2.13	1.43	4.00
	Johnny Vaughan	3.95	1.80	2.65	4.00
	Jo O'Meara	4.48	2.00	3.00	3.62
	Kate Winslett	3.95	2.30	2.75	4.00
	Leonardo Di Caprio	3.81	2.55	3.00	4.38
	Mel C	3.95	2.00	2.65	3.14
	Michelle Pfeiffer	3.52	2.70	2.20	3.67
	Reece Witherspoon	4.60	2.55	2.60	4.24
	Ricky Gervais	4.95	2.35	2.85	3.90
	Ross Kemp	3.62	2.25	2.00	4.71
	Simon Cowell	4.95	2.30	2.80	3.86
	Susannah Constantine	4.90	2.65	3.00	3.30
	Trisha Goddard	4.43	2.60	3.05	4.14

Appendix 4. *Stimuli from Experiment 3.*

		Ratings			
Condition	Stimulus	AoA	Distinctiveness	Frequency	Likeness
Early acquired distinctive	Billy Connelly	3.29	3.67	2.45	3.36
	Cilla Black	2.38	3.37	2.55	4.69
	David Jason	2.71	3.11	2.75	4.00
	Des Lynam	2.90	3.04	2.40	4.27
	Gary Lineaker	2.81	3.15	2.75	3.38
	Goldie Hawn	3.24	3.19	2.15	3.30
	Joanna Lumley	3.38	3.22	2.15	2.75
	John Travolta	2.95	3.74	2.45	4.24
	Julia Roberts	3.05	3.04	3.05	3.92
	Kylie Minogue	2.33	3.22	3.95	2.46
	Lenny Henry	3.24	2.78	2.35	3.54
	Pamela Anderson	3.19	2.93	2.30	4.04
	Prince Charles	2.14	3.70	3.15	3.85
	Princess Diana	2.26	2.77	2.61	3.48
	Rod Stewart	3.24	3.74	2.00	3.86
	Sean Connery	2.71	3.04	2.40	3.12
	Terry Wogan	2.76	2.85	2.25	3.77
	Tom Jones	3.00	3.27	2.35	2.16
Late acquired distinctive	Billy Piper	4.00	2.75	1.70	3.90
	Brad Pitt	3.62	3.26	3.40	3.31
	Chris Evans	3.57	4.26	1.90	4.27
	Eminem	4.62	2.75	3.79	3.24
	Hugh Laurie	3.67	4.30	1.95	3.63
	Jennifer Lopez	4.43	3.35	4.00	3.67
	Jeremy Clarkson	3.81	3.59	2.64	3.80
	Judi Dench	3.48	3.07	2.65	3.92
	Liv Tyler	4.35	3.11	2.70	3.61
	Martin Clunes	3.62	3.89	2.35	4.42
	Meg Ryan	3.48	2.78	2.25	3.00
	Michael Caine	3.48	3.27	2.25	3.79
	Naomi Campbell	3.52	2.89	2.45	3.33
	Ricky Martin	4.19	3.20	1.85	3.81
	Russell Crowe	4.67	2.85	2.75	4.05
	Sophie Ellis Bextor	4.71	4.35	2.65	4.00
	Stephen Fry	3.81	3.00	2.73	4.08
	Vinnie Jones	3.90	3.00	2.47	4.48
Early acquired typical	Alan Shearer	3.22	1.70	2.26	3.32
	Andi Peters	2.29	2.08	1.21	3.88
	Anthea Turner	2.38	2.15	1.85	4.48
	Carol Vorderman	3.10	2.31	3.05	3.40
	Cheryl Baker	2.33	1.93	1.32	3.57
	Danni Minogue	2.95	2.70	3.40	1.42

Appendix 4. *Stimuli from Experiment 3 – Continued.*

Condition	Stimulus	Ratings			
		AoA	Distinctiveness	Frequency	Likeness
Early acquired typical	Harrison Ford	2.71	2.26	2.40	4.08
	John Leslie	2.38	2.15	2.50	3.38
	Keith Chegwin	3.35	2.08	2.14	3.29
	Les Dennis	3.29	2.44	2.20	3.92
	Linda Robson	2.33	2.00	1.70	3.68
	Michael Barrymore	2.81	2.59	2.42	3.64
	Pauline Quirke	2.38	2.37	2.00	3.71
	Ronan Keating	3.41	2.69	2.57	3.72
	Todd Carty	3.25	2.58	2.91	3.52
	Victoria Wood	2.95	2.56	1.58	3.82
	Wendy Richard	2.67	2.38	3.45	2.56
	Zoe Ball	3.43	2.22	2.95	3.77
Late acquired typical	Angus Deayton	3.76	2.70	2.40	4.08
	Anna Kournikova	4.29	2.25	2.50	2.90
	Cameron Diaz	3.95	2.33	2.90	3.17
	Ewan McGregor	3.95	2.59	2.75	3.12
	Harry Enfield	3.62	2.44	2.45	2.88
	Jill Dando	3.83	2.08	1.43	3.88
	Johnny Vaughan	3.95	1.88	2.65	3.21
	Jonathon Ross	3.90	2.67	3.05	3.77
	Jo O'Meara	4.48	2.00	3.00	3.62
	Kate Winslett	3.95	2.07	2.75	2.69
	Leonardo Di Caprio	3.81	2.63	3.00	4.23
	Mel C	3.95	2.30	2.65	2.96
	Michelle Pfeiffer	3.52	2.44	2.20	2.36
	Reece Witherspoon	4.60	2.55	2.60	4.24
	Ricky Gervais	4.95	2.35	2.85	3.90
	Ross Kemp	3.62	2.26	2.00	4.13
	Simon Cowell	4.95	2.30	2.80	3.86
	Trisha Goddard	4.43	2.60	3.05	4.14

Appendix 5. *Stimuli from Experiments 5 and 6.*

		Ratings				
Conditior	Stimulus	AoA	Dist.	Frequency	Likeness	Familiarity
Early acquired	Anthea Turner	2.38	2.35	1.85	4.29	4.32
	Barbara Windsor	3.14	2.95	3.70	4.48	4.45
	Carol Vorderman	3.10	2.60	3.05	3.33	4.64
	Cher	2.71	3.80	2.05	4.38	4.59
	Cilla Black	2.38	3.50	2.55	4.62	4.64
	Danni Minogue	2.95	2.60	3.40	3.81	3.73
	Joanna Lumley	3.38	3.65	2.15	4.20	4.41
	Julia Roberts	3.05	3.35	3.05	4.38	4.91
	Kylie Minogue	2.33	3.30	3.95	4.29	4.86
	Leslie Joseph	3.09	3.70	2.05	3.95	3.86
	Linda Robson	2.33	1.80	1.70	3.76	3.95
	Madonna	2.62	3.25	3.80	3.67	4.95
	Oprah Winfrey	3.38	2.90	2.10	3.90	4.00
	Princess Diana	2.26	3.04	2.61	3.74	4.95
	Queen Mother	1.74	2.87	1.91	3.30	4.14
	Wendy Richard	2.67	2.15	3.45	4.24	3.82
	Zoe Ball	3.43	2.55	2.95	4.19	4.36
	Alan Shearer	3.22	2.43	2.26	3.64	4.23
	Arnold Schwarzenegger	2.67	3.90	2.05	3.81	4.82
	Gary Lineaker	2.81	3.20	2.75	4.10	4.77
	Harrison Ford	2.71	2.45	2.40	4.05	4.05
	John Leslie	2.38	1.65	2.50	4.29	4.55
	John Travolta	2.95	2.90	2.45	4.14	4.77
	Keith Chegwin	3.35	2.39	2.14	3.89	4.00
	Les Dennis	3.29	1.70	2.20	3.86	3.91
	Michael Barrymore	2.81	1.90	2.42	4.05	4.27
	Michael Jackson	2.14	4.80	3.65	4.62	4.95
	Nicholas Lyndhurst	2.91	3.78	3.17	4.04	4.55
	Prince Charles	2.14	3.65	3.15	3.95	4.86
	Rod Stewart	3.24	3.60	2.00	4.24	3.77
	Ronan Keating	3.41	1.83	2.57	3.00	4.32
	Rowan Atkinson	2.38	3.90	2.20	3.67	4.73
	Terry Wogan	2.76	2.95	2.25	3.95	4.55
	Todd Carty	3.25	2.52	2.91	4.21	3.68
Late acquired	Billy Piper	4.00	2.75	1.70	3.90	4.00
	Calista Flockhart	4.14	3.25	3.00	4.33	4.64
	Cameron Diaz	3.95	2.65	2.90	4.00	4.82
	Denise Van Outen	3.81	2.45	3.25	4.05	4.41
	Halle Berry	4.71	3.55	3.05	4.52	4.50
	Jade Goody	5.00	3.55	3.20	4.24	4.09
	Jill Dando	3.83	2.13	1.43	4.00	4.00

Appendix 5. *Stimuli from Experiments 5 and 6 – Continued.*

		Ratings				
Conditior	Stimulus	AoA	Dist.	Frequency	Likeness	Familiarity
Late	Judi Dench	3.48	3.09	2.65	4.48	4.77
acquired	Kate Winslett	3.95	2.30	2.75	4.00	4.59
	Lisa Riley	3.95	3.45	2.80	4.05	4.05
	Liv Tyler	4.35	3.15	2.70	4.24	4.18
	Meg Ryan	3.48	2.80	2.25	3.25	4.23
	Michelle Pfeiffer	3.52	2.70	2.20	3.67	3.95
	Reece Witherspoon	4.60	2.55	2.60	4.24	4.68
	Sarah Michelle Gellar	4.19	2.75	3.40	4.10	4.77
	Sophie Ellis Bextor	4.71	4.35	2.65	4.00	4.68
	Ulrika Jonsson	3.48	2.05	3.05	3.62	4.45
	Chris Evans	3.57	3.65	1.90	3.81	4.68
	Darius	4.90	2.30	3.15	4.19	4.32
	Gareth Gates	5.00	2.80	3.70	4.52	4.77
	Harry Enfield	3.62	2.55	2.45	3.38	4.45
	Hugh Laurie	3.67	4.13	1.95	3.61	3.95
	Jeremy Clarkson	3.81	3.57	2.64	3.86	4.32
	Johnny Vaughan	3.95	1.80	2.65	4.00	4.36
	Lawrence Llewellyn Bowen	4.29	3.00	2.90	3.76	4.41
	Leonardo Di Caprio	3.81	2.55	3.00	4.38	4.82
	Martin Clunes	3.62	3.90	2.35	4.10	4.59
	Michael Owen	4.24	1.60	3.15	4.14	4.45
	Pierce Brosnan	3.95	2.75	2.55	4.24	4.73
	Ricky Martin	4.19	3.20	1.85	3.81	4.41
	Ross Kemp	3.62	2.25	2.00	4.71	4.14
	Stephen Fry	3.81	3.91	2.73	3.38	4.09
	Vinnie Jones	3.90	3.40	2.47	4.20	4.32
	Will Young	5.00	3.20	3.10	3.95	4.32

Appendix 6. *Stimuli from Experiments 7 and 8.*

		Ratings			
Conditior	Stimulus	AoA	Dist.	Frequency	Likeness
Early acquired	Anthea Turner	2.38	2.15	1.85	4.48
	Barbara Windsor	3.14	3.41	3.70	2.72
	Carol Vorderman	3.10	2.31	3.05	3.40
	Cilla Black	2.38	3.37	2.55	4.69
	Julia Roberts	3.05	3.04	3.05	3.92
	Kylie Minogue	2.33	3.22	3.95	2.46
	Lesley Joseph	3.09	2.78	2.05	3.52
	Linda Robson	2.33	2.00	1.70	3.68
	Madonna	2.62	3.11	3.80	2.52
	Margaret Thatcher	2.33	3.63	1.95	4.15
	Oprah Winfrey	3.38	3.33	2.10	4.27
	Pauline Quirke	2.38	2.37	2.00	3.71
	Princess Diana	2.26	2.77	2.61	3.48
	Wendy Richard	2.67	2.38	3.45	2.56
	Zoe Ball	3.43	2.22	2.95	3.77
	Alan Shearer	3.22	1.70	2.26	3.32
	Arnold Schwarzenegger	2.67	3.23	2.05	3.76
	Billy Connolly	3.29	3.67	2.45	3.36
	Gary Lineaker	2.81	3.15	2.75	3.38
	Harrison Ford	2.71	2.26	2.40	4.08
	John Leslie	2.38	2.15	2.50	3.38
	Keith Chegwin	3.35	2.08	2.14	3.29
	Michael Barrymore	2.81	2.59	2.42	3.64
	Michael Jackson	2.14	4.81	3.65	4.32
	Nicholas Lyndhurst	2.91	3.26	3.17	4.46
	Prince Charles	2.14	3.70	3.15	3.85
	Robin Williams	2.90	2.96	2.50	2.24
	Ronan Keating	3.41	2.69	2.57	3.72
	Terry Wogan	2.76	2.85	2.25	3.77
	Todd Carty	3.25	2.58	2.91	3.52
Late acquired	Anna Kournikova	4.29	2.25	2.50	2.90
	Britney Spears	4.33	3.05	3.70	4.52
	Cameron Diaz	3.95	2.33	2.90	3.17
	Christina Aguilera	4.52	3.45	3.90	4.10
	Holly Valance	4.76	2.50	3.20	4.14
	Jennifer Lopez	4.43	3.35	4.00	3.67
	Jill Dando	3.83	2.08	1.43	3.88
	Judi Dench	3.48	3.07	2.65	3.92
	Lisa Riley	3.95	3.15	2.80	4.44
	Liv Tyler	4.35	3.11	2.70	3.61
	Meg Ryan	3.48	2.78	2.25	3.00

Appendix 6. *Stimuli from Experiments 7 and 8 – Continued.*

		Ratings			
Conditior	Stimulus	AoA	Dist.	Frequency	Likeness
Late acquired	Mel C	3.95	2.30	2.65	2.96
	Michelle Pfeiffer	3.52	2.44	2.20	2.36
	Naomi Campbell	3.52	2.89	2.45	3.33
	Trisha Goddard	4.43	2.60	3.05	4.14
	Angus Deayton	3.76	2.70	2.40	4.08
	Chris Evans	3.57	4.26	1.90	4.27
	Eminem	4.62	2.75	3.79	3.24
	Ewan McGregor	3.95	2.59	2.75	3.12
	Harry Enfield	3.62	2.44	2.45	2.88
	Hugh Laurie	3.67	4.30	1.95	3.63
	Jamie Oliver	4.48	2.40	3.40	3.95
	Jeremy Clarkson	3.81	3.59	2.64	3.80
	Johnny Vaughan	3.95	1.88	2.65	3.21
	Jonathon Ross	3.90	2.67	3.05	3.77
	Martin Clunes	3.62	3.89	2.35	4.42
	Ross Kemp	3.62	2.26	2.00	4.13
	Shane Ritchie	3.70	2.26	3.95	2.95
	Simon Cowell	4.95	2.30	2.80	3.86
	Stephen Fry	3.81	3.00	2.73	4.08

Appendix 7. *Stimuli from Experiments 9 and 10.*

Condition	Stimulus	Ratings				
		AoA	Dist.	Frequency	Likeness	Whole face gender typicality
Early High Gender Typicality	Alan Shearer	3.22	2.43	2.26	3.64	4.68
	Arnold Schwarzenegger	2.67	3.90	2.05	3.81	5.59
	Gary Lineaker	2.81	3.20	2.75	4.10	4.36
	Harrison Ford	2.71	2.45	2.40	4.05	4.77
	John Travolta	2.95	2.90	2.45	4.14	5.23
	Lenny Henry	3.24	2.85	2.35	3.67	5.05
	Michael Barrymore	2.81	1.90	2.42	4.05	4.45
	Prince Charles	2.14	3.65	3.15	3.95	4.23
	Rowan Atkinson	2.38	3.90	2.20	3.67	5.14
	Todd Carty	3.25	2.52	2.91	4.21	4.59
Early Low Gender Typicality	David Bowie	2.76	3.70	1.80	3.43	3.18
	John Leslie	2.38	1.65	2.50	4.29	3.50
	Keith Chegwin	3.35	2.39	2.14	3.89	3.68
	Les Dennis	3.29	1.70	2.20	3.86	3.45
	Michael Jackson	2.14	4.80	3.65	4.62	1.32
	Nicholas Lyndhurst	2.91	3.78	3.17	4.04	3.91
	Rod Stewart	3.24	3.60	2.00	4.24	3.68
	Ronan Keating	3.41	1.83	2.57	3.00	3.18
	Terry Wogan	2.76	2.95	2.25	3.95	3.91
	Tom Jones	3.00	3.30	2.35	4.19	3.64
Late high Gender Typicality	Angus Deayton	3.76	2.65	2.40	3.62	4.36
	Chris Evans	3.57	3.65	1.90	3.81	4.41
	Ewan McGregor	3.95	2.25	2.75	3.52	4.59
	Jeremy Clarkson	3.81	3.57	2.64	3.86	4.41
	Martin Clunes	3.62	3.90	2.35	4.10	4.27
	Michael Owen	4.24	1.60	3.15	4.14	4.27
	Pierce Brosnan	3.95	2.75	2.55	4.24	5.59
	Ross Kemp	3.62	2.25	2.00	4.71	4.91
	Simon Cowell	4.95	2.30	2.80	3.86	5.05
	Vinnie Jones	3.90	3.40	2.47	4.20	5.64
Late Low Gender Typicality	Darius	4.90	2.30	3.15	4.19	3.27
	Gareth Gates	5.00	2.80	3.70	4.52	2.09
	Harry Enfield	3.62	2.55	2.45	3.38	4.23
	Hugh Laurie	3.67	4.13	1.95	3.61	4.23
	Johnny Vaughan	3.95	1.80	2.65	4.00	4.18
	Jonathon Ross	3.90	2.50	3.05	3.57	3.14
	Lawrence Llewellyn Bowen	4.29	3.00	2.90	3.76	2.05
	Ricky Martin	4.19	3.20	1.85	3.81	3.64
	Stephen Fry	3.81	3.91	2.73	3.38	3.77
	Will Young	5.00	3.20	3.10	3.95	2.91