# Memory Processes in familiar voice recognition.

Ljubica Damjanovic

A thesis submitted for the degree of Doctor of Philosophy in Psychology

Department of Psychology

University of Essex

October 2004

# Abstract.

Research suggests that there are cross-modal differences in the retrieval of person-specific information from faces and voices. Specifically, Hanley, Smith and Hadfield (1998) have shown that we are much more likely to recall the occupation and name from a celebrity's face that we deem familiar than from their voice. In person-recognition terms, a failure to 'place' a person is known as the familiar-only error. The aim of this thesis was to elucidate the cognitive factors that may be contributing to the large number of familiar-only responses that occur in voice recognition. Chapter 3 extended Hanley et al's (1998) initial observation of a familiar-only bias in voice recognition with a new selection of stimuli (Experiments 1a and 1d). The hypothesis that participants were poorly calibrated in their familiarity judgments with faces and voices was tested using a within-subjects design (Experiments 1b, 1c and 1e). Even under these testing conditions participants found the retrieval of person-specific information to voices deemed familiar particularly problematic. However, Chapter 4 failed to produce a modality difference in occupational and name recall to familiar faces and voices when participants were required to make confidence judgments. Chapter 4 also used confidence judgments to generate ROCs and z-ROCs. This method raised the possibility that face recognition may entail a memory component that relies on recall whereas voice recognition may entail a familiarity process. However, methodological aspects of the study were not sensitive enough to provide strong support for this conclusion. Chapter 5 showed that applying a blur filter to faces can successfully reduce overall recognition performance to a comparable level with voices. This new blurred face condition was used as a matched control with voices in Chapter 6. The critical finding was that even though blurred faces and voices did not differ in overall performance, they differed significantly in terms of the memory processes that they rely on. Specifically, blurred faces are associated more strongly with remembering, whereas voices are more strongly associated with knowing (Experiment 4a and 4c). This difference can not be attributed to differences in subjective confidence (Experiment 4b). The results suggest that the cross-modal difference in person-identification can be understood in terms of the qualitatively different recognition memory processes that each modality relies on.

## Acknowledgements.

Personally, my gratitude goes to Professor Rick Hanley for inspiring my interest in face and voice recognition. To my second supervisor, Dr Steve Avons, whose help more often than not exceeded the duties of a second supervisor. To Professor Elaine Fox and Dr Debi Roberson who provided indirect support by introducing me to related areas of research to my PhD. To the Department of Psychology's technical support team, but particularly to Roger Deeble, whose vast expertise and insight helped me overcome even the most seemingly insoluble problems.

Personally, my gratitude goes to the Damjanovic family, the Simonovic family, the Athanasopoulos family and Mrs Maria Worrall for their love and understanding. Their support has helped me morally and spiritually.

Most of all, I would like to express infinite gratitude to my partner, Panos Athanasopoulos. Thanks for your time, patience and understanding and helping me to make my dreams come true.

## Contents.

Abstract……page i

Acknowledgements……iii

Index of figures and tables……viii

**Chapter 1: Introduction**……1

1.1: What's in a voice?……1

1.2: Voice recognition – how easy is it?……5

1.3: Do faces and voices go together?……10

1.4: The voice area……15

1.5: Voice as evidence……20

1.6: Person recognition memory models – boxes and arrows……23

1.7: The Cognitive system……34

1.8: Burton, Bruce and Johnston – an interactive activation and competition (IAC) model……35

1.9: Aims of thesis……40

**Chapter 2: Selection of stimuli**……44

Part 1 – selection process……44

Part 2a: Faces – controlling for semantic cues……45

Part 2b: Voices – controlling for semantic cues……48

2.2: Final selection……50

**Chapter 3: The relationship between familiarity and person Identification. Is it the same for faces and voices?**……51

3.1: Introduction……51

3.2: Priming in person recognition memory……52

3.3: The role of the PINs in face and voice recognition..........................56

3.4: The Mixed question effect..................................................................62

3.5: Rationale..........................................................................................66

Experiment 1a........................................................................................68

Method..................................................................................................68

Results..................................................................................................71

Discussion.............................................................................................77

Experiment 1b........................................................................................79

Method..................................................................................................79

Results..................................................................................................81

Discussion.............................................................................................85

Experiment 1c........................................................................................87

Method..................................................................................................87

Results..................................................................................................89

Discussion.............................................................................................94

Experiment 1d........................................................................................97

Method..................................................................................................97

Results..................................................................................................101

Discussion.............................................................................................105

Experiment 1e........................................................................................107

Method..................................................................................................107

Results..................................................................................................109

Discussion.............................................................................................113

General Discussion.................................................................................114

**Chapter 4: Confidence judgments in person recognition memory**........123

4.1: Introduction......123

4.2: Confidence judgments in person recognition memory......123

4.3: Confidence and qualitative processes in person recognition memory....134

4.4: Evidence for a dual process account obtained through confidence judgments......135

4.5: Rationale......140

*Experiment 2*......142

*Method*......142

*Results*......145

General Discussion......156

**Chapter 5: Matching face and voice recognition**......166

5.1: Introduction......166

5.2: Rationale......166

a.) Inverted faces......167

b.) Blurred faces......169

c.) Negative faces......170

Experiment 3......172

Method......172

Results......179

General Discussion......189

**Chapter 6: The R and K paradigm. Can it be applied to the Recognition of familiar faces and voices?**......195

6.1: Introduction......195

6.2: The R and K paradigm......196

6.3: Remembering and Knowing – a quantitative explanation?......200

6.4: The R/K paradigm and person recognition....................................204

6.5: Rationale..........................................................................................206

Experiment 4a.........................................................................................211

Method......................................................................................................211

Results......................................................................................................217

Discussion................................................................................................227

Experiment 4b.........................................................................................231

Method......................................................................................................231

Results......................................................................................................233

Discussion................................................................................................235

Experiment 4c.........................................................................................237

Method......................................................................................................237

Results......................................................................................................239

Discussion................................................................................................245

General Discussion..................................................................................247

**Chapter 7: Conclusions**......................................................................258

7.1: Conclusions.......................................................................................258

7.2: Final Comments...............................................................................265

**References**...........................................................................................267

APPENDIX 1..............................................................................................290

APPENDIX 2..............................................................................................291

APPENDIX 3..............................................................................................292

APPENDIX 4..............................................................................................293

APPENDIX 5..............................................................................................295

## Index of figures and tables.

*Figures.*

**Chapter 1:**

1.6a: Bruce and Young's (1986) functional model of Face recognition.......24

1.6b: Simplified version of Bruce and Young's face recognition route with proposed voice recognition route by Belin, Fecteau and Bédard..........25

1.8a: The IAC Model (Burton and Bruce, 1993).................................36

**Chapter 2:**

2.1a: Original version of John Major clip with clothing cues...................47

2.1b: John Major clip with clothing cues removed with cropping tool from Final Cut Pro............................................................47

**Chapter 3:**

3.1a: Mean number of responses at each level of familiarity for faces and voices..............................................................74

3.1b: Mean proportion of occupations recalled at each level of familiarity for faces and voices.................................................76

3.1c: Mean proportion of names recalled at each level of familiarity for faces and voices.................................................77

3.2a: Mean number of responses at each level of familiarity for faces and voices...............................................................83

3.2b: Mean proportion of occupations recalled at each level of familiarity for faces and voices.................................................84

3.2c: Mean proportion of names recalled at each level of familiarity for faces and voices.................................................85

3.3a: Mean number of responses at each level of familiarity for faces and voices........91

3.3b: Mean proportion of occupations recalled at each level of familiarity for faces and voices........92

3.3c: Mean proportion of names recalled at each level of familiarity for faces and voices........93

3.4a: Mean number of responses at each level of familiarity for faces and voices........103

3.4b: Mean proportion of occupations recalled at each level of familiarity for faces and voices........104

3.4c:Mean proportion of names recalled at each level of familiarity for faces and voices........105

3.5a: Mean number of responses at each level of familiarity for faces and voices........111

3.5b: Mean proportion of occupations recalled at each level of familiarity for faces and voices........112

3.5c: Mean proportion of names recalled at each level of familiarity for faces and voices........113

**Chapter 4:**

4.2a: Overconfidence bias in voice recognition........131

4.4a: Hypothetical ROCs and z-ROCs........137

4.1a: Proportion of Hits (OLD) made at each level of confidence for faces and voices........148

4.1b: Proportion of occupations recalled at each level of confidence for faces and voices........149

4.1c: Proportion of names recalled at each level of confidence for faces and voices............150

4.1d: Average ROCs and z-ROCs for the face condition...........154

4.1e: Average ROCs and z-ROCs for the voice condition...........155

**Chapter 5:**

5a: Chris Tarrant. Image rotated through 180° using Final Cut Pro...........173

5b: Chris Tarrant. Blur filter at level 13 using Final Cut Pro...........174

5c: Chris Tarrant. Photographic negative using Final Cut Pro...........175

5.1a: Average ROCs and z-ROCs for the inverted face condition...........183

5.1b: Average ROCs and z-ROCs for the blurred face condition...........185

5.1c: Average ROCs and z-ROCs for the negative face condition...........188

**Chapter 6:**

6.1a: Proportion of correct Remember, Know and Guess responses for each modality condition...........222

6.1b: Proportion of occupations retrieved for Remember, Know and Guess responses as a function of modality...........227

6.2a: Proportion correct for very sure, less sure and guess responses for the blurred face and voice conditions...........235

6.3a: Proportion of correct Remember, Know and Guess responses in each condition...........241

6.3b: Proportion of occupations retrieved for Remember, Know and Guess responses as a function of modality...........245

*Tables.*

**Chapter 3:**

3.1a: Means and standard deviations for the face and voice conditions.......71

3.2a: Means and standard deviations for the face and voice conditions.......81

3.3a: Means and standard deviations for the face and voice conditions.......89

3.4a: Means and standard deviations for the face and voice conditions.......101

3.5a: Means and standard deviations for the face and voice conditions.......109

**Chapter 4:**

4.1a: Means and standard deviations for the face and voice conditions.......145

**Chapter 5:**

5.1a: Performance measures for inverted, blurred and negative faces. Standard face and voice means are taken from Chapter 4........................179

**Chapter 6:**

6.1a: Mean overall hits, false alarm rates and their corresponding A' (sensitivity) and B"D (bias) measures for the standard face, blurred Face and voice conditions.........................................................218

6.1b: Mean proportion of correct (hits) and incorrect (false alarms) remember, know and guess responses for the standard face, blurred face and voice conditions.............................................................220

6.1c: Mean sensitivity (A') and criterion measures (B"D) calculated separately for remember, know and guess responses...........................223

6.1d: Mean sensitivity (A') and criterion measures (B"D) for remember only and remember plus know data for the three modality conditions........224

6.1e: Mean proportion of occupational descriptions given to remember, know and guess responses across the three modality conditions..............226

6.2a: Mean overall hits, false alarm rates and their corresponding A' (sensitivity) and B"D (bias) measures for the blurred face and voice conditions..........233

6.2b: Mean proportion of correct and incorrect very sure, less sure and guess responses for the blurred face and voice conditions........................234

6.3a: Mean overall hits, false alarm rates and their corresponding A'(sensitivity) and B"D (bias) measures for the blurred face and voice conditions...........239

6.3b: Mean proportion of correct and incorrect remember, know and guess responses for the blurred face and voice conditions........................240

6.3c: Mean sensitivity (A') and criterion measures (B"D) calculated separately for remember, know and guess responses.............................242

6.3d: Mean sensitivity (A') and criterion measures (B"D) for remember only and remember plus know data for the blurred face and voice condition......243

6.3e: Mean proportion of correct occupational descriptions given to remember, know and guess responses across the two modality conditions...244

# Chapter 1: Introduction — Memory processes in familiar voice recognition.

Throughout the animal kingdom the ability to recognize other conspecifics is an integral part of adaptive behaviour (Insley, 2000, Johnson, 1992). Contemporary models of human person recognition (Bruce and Young, 1986, Burton, Bruce and Johnston, 1990) have proposed different perceptual and cognitive stages that allow us to encode unfamiliar faces and recognize familiar people. The validity of these models has largely been dependent on the research findings from face recognition. Undoubtedly, recognizing people is not just dependent on the visual domain. The ability to recognize other people from hearing their voice is also a source for identity information. Therefore these models have proposed (albeit implicitly) an auditory route, that runs in parallel to the visual one, to capture this notion of 'person' rather than face recognition. These models and the empirical studies that test the assumptions of these models shall be provided in a separate section of this chapter.

The central aim of this thesis is to look at the memory processes involved in familiar voice recognition and investigate whether these processes are the same as those that govern familiar face recognition. The ways in which these processes will be measured are outlined in the final section of this chapter. For now, a summary on the research on voice recognition shall be presented. These summaries will be structured around psychological, perceptual and applied issues relevant to voice recognition research.

## 1.1: What's in a voice?

The obvious critical role of the human voice is to carry the speech signal required for communication. However, it is also the carrier of identity information. Inarguably, this makes the human voice the most important sound in our auditory environment with some researchers referring to it as the 'auditory face' (Belin, Fecteau and Bédard, 2004). What makes the study of voice recognition particularly interesting is that long before language became a part of our central form of communication, the evolutionary history of the human voice shows that our ancestors relied heavily on non-linguistic properties of the voice such as cries and other vocalizations to survive and communicate in their environment (Hauser, 1996).

According to Sapir (1927) the speech signal involves five properties; voice, speech dynamics, pronunciation, vocabulary style and style of connected utterance. The focus of this thesis will be on the first property, however that is not to say that the other properties will be excluded, as collectively they all play a role in what we perceive as a 'voice'. The production of the speech signal is a rather mechanical process which begins when the air from the lungs passes along the trachea and into the larynx. The air then passes through two muscular folds called the vocal cords. Past the larynx is the area known as the vocal tract, which contains some of the articulators that modify the speech signal, such as the tongue and lips (Ladefoged, 2001).

The rate at which the vocal cords open and close is referred to as frequency and it is measured in Hertz (Hz). The frequency range of the normal speech signal is 80 — 8000 Hertz (Hz) and its intensity or perceived loudness ranges from 60 — 70 decibels (dB). The average fundamental frequency of vocal cord vibration, which is associated with the subjective quality of pitch, is between 180 and 300 Hz for females and 90 and 140 Hz for males (Meuwly, 2000). The human ear can perceive frequencies ranging from 30 to 20,000 Hz (Toates, 1998). It is particularly sensitive to frequencies that fall within 500 and 4000 Hz, which is convenient given the frequency range of the speech signal.

The message carried in the speech signal is perceived through a five stage process. These processes are ordered as peripheral auditory analysis, central auditory analysis, acoustic-phonetic analysis, phonological analysis and higher order analysis which can be broken down into lexical, syntactic and semantic forms (Meuwly, 2000).

The human voice and the speech signal that it contains are inextricably linked. For example, the rise in pitch at the end of sentence can be used as a linguistic marker to denote that a statement is a question, such as 'John is going'. Furthermore, there are certain languages that rely on pitch differences to convey semantic as well as syntactic properties of a language. For example, in Mandarin Chinese the sound 'ma' can either mean mother, hemp, horse or cold depending on where the pitch rises and/or falls (Ball and Rahilly, 1999).

By the same token, changes in pitch coupled with changes in loudness (Scherer, 1981) can be used to convey different emotional expressions. Just as human facial expressions are characterized in terms of the contortion of various facial muscles or action units (Ekman and Freisen, 1975, Waters, 1992), so are voices characterized by changes in pitch and loudness. For example, the acoustic parameters of happiness include large pitch variations along with moderate variations in amplitude (Scherer, 1981). As a consequence, recent modifications to psychological assessments of emotional perception have now incorporated vocal expressions along with facial ones (for example see the Comprehensive Affect Testing System by Froming, Levy and Ekman, 2003).

The review so far has looked at perceptual and linguistic properties of a voice. What needs to be established now is how identity information can be extracted. It is important to consider at this point, what is defined as identity. A definition for the current purposes of this chapter will be taken from Bruce and Young (1986). Bruce and Young (1986) reported two types of identity codes, these are termed as visually derived semantic codes and identity-specific semantic codes. In face recognition terms, the former type of code is created by physical properties of a face such as age, sex, gender and expression. Non-physical characteristics can also be formed, such as judging the intelligence or honesty of a face. Identity-specific semantic codes convey descriptions of the person's occupation and where they are usually encountered and so on. Bruce and Young (1986) argue that the retrieval of identity-specific codes gives us a sense of 'knowing' to whom the face belongs to. The way in which this code is retrieved will be covered in greater detail in section 1.6. For now, research relating to age and sex judgements or 'acoustically' derived semantic codes in voice recognition shall be briefly summarized.

Ptacek and Sander (1966) observed near perfect performance when listeners had to make a two-category age judgments to voices based on the reading of a complete passage from a text. Performance continued to be above chance even when judgments were required for single vowel sounds. Shipp and Hollien (1969) extended this area of research by showing that listeners could make finer age discriminations (20-30yrs, 30-40 yrs etc) than the simple old-young distinctions obtained in previous studies.

The ability to identify the sex of the speaker is a robust skill and can survive various disguise attempts such as talking in whispers (Schwartz and Rine, 1968). Variation in vocal characteristics between sexes is largely a result of physiological differences. The pharyngeal cavity is shorter in females which affects the variation in vowel formant frequencies. This is particularly marked for extreme tongue position vowels, such as /i/ as in 'team' (Coleman, 1976). Other characteristics that can apparently be derived from voices are more controversial in their nature and include judging the social status of the speaker (Sapir, 1927) and even judging the sexual attractiveness of the speaker (Collins, 2000).

Whilst these studies demonstrate how readily we extract acoustically derived semantic codes from voices (i.e; age, sex and attractiveness), a greater understanding is needed of how we achieve a sense of voice recognition. In other words, how well we can recognize a person's voice that we previously encountered. The next section moves on from the social characteristics that can be identified from a person's voice to the area of voice recognition in general.

### 1.2: Voice recognition — how easy is it?

McGehee (1937, 1944) is often credited with carrying out the first voice recognition experiment in the experimental laboratory. The experiment was inspired by various court cases that were prominent at that time. Additional court cases which incorporate voice information as forensic evidence can be found in section 1.5 of this Chapter. One particular case that served as the rationale behind McGehee's research was the case of United States v. Hauptmann (cited in Saslove and Yarmey, 1980). This controversial case involved Colonel Charles Lindburgh claiming in court that he was able to recognize the defendant's voice as that of his son's kidnapper. This claim was based on two words that the defendant spoke three years earlier.

McGehee (1937) wanted to evaluate the validity of such legal evidence, especially since no prior experimental investigations were conducted on voice recognition accuracy. Participants were required to listen to a reading made by the target speaker, who was concealed from the participants. After a delay of either 2 days, 2 weeks or 5 months, the participants returned to the laboratory where they had to select the target speaker from four other speakers. The accuracy rates were 80% for the 2 day delay, after which accuracy dropped rapidly reaching 13% in 5 months. Accuracy was impaired even further, if participants were exposed to several voices during the initial listening stage. Additional research that has aimed to clarify the effects of delay on identification accuracy has shown that accuracy performance as measured in hits does not differ for voices tested after 5 minutes, 24 hours or 1 week. However, the number of errors or false alarms significantly increased over these periods, particularly for voices that were initially heard with short (18s and 38s) exposure durations (Yarmey and Matthys, 1992 cited in Yarmey, 1995).

Whilst this study by McGehee (1937) is considered to be a pioneering landmark by contemporary researchers in the voice recognition field, the conclusions from this study have been criticized on a number of methodological grounds. As Saslove and Yarmey (1980) pointed out, one of the design problems with McGehee's original study was that participants were not randomly allocated to the various delay conditions. The study was also problematic from a procedural viewpoint, because different target voices were used in each test period.

A theoretical concern was raised by Van Lancker, Kreiman and Emmorey (1985) who argued that the study conducted by McGehee (1937) and subsequent researchers (Voiers, 1964) have investigated the cognitive abilities of voice *discrimination*, a term which is often used synonymously with voice recognition. Studies such as the one conducted by McGehee (1937) included a voice set which were unknown to the listener. Van Lancker et al (1985) argue that the cognitive process underlying the ability to recognize a familiar voice, such as a friend's voice, is a very different process from discriminating among unfamiliar voices. Van Lancker and Kreiman (1987) explained the differences between voice recognition and voice discrimination by suggesting a Gestalt type process associated with recognition, whereas a more featural process being associated with discrimination. The Gestalt aspect of voice recognition involves a pattern recognition process in which the unique vocal qualities of the voice pattern is linked to a name or person. Voice discrimination also involves an overall pattern recognition process, however this is primarily driven by a featural analysis of the various cues (pitch and speaking rate, for example) in the acoustic signal and judging the degree to which two voices differ along these acoustic dimensions.

The contrast in use of unfamiliar vs. familiar voices as experimental stimuli has generally reflected the rationale between two schools of thought. For example, researchers from the forensic field are especially interested in recreating real-life witnessing situations. As such, participants are required to select a voice which they had encountered previously under non-intentional encoding conditions from a line-up (Yarmey, Yarmey and Yarmey, 1994). When participants are informed of the voice identification test, overall selection of the target improves significantly relative to a condition in which participants are not informed about the subsequent test (Salsove and Yarmey, 1980).

The cognitive school of thought tends to focus on how readily we can recognize voices and retrieve person-specific information to them. As a result, personally familiar (Yarmey, Yarmey, Yarmey and Parliament, 2001) or publicly familiar voices are used (Van Lancker, Kreiman and Emmorey ,1985, Schweinberger, 2001). This following section provides an outline of the research findings associated with familiar voice recognition.

Studies that have used familiar voices as test stimuli seem to indicate that listeners are generally quite good at recognizing familiar voices. In one particular study Ladefoged and Ladefoged (1980) used 29 voices from family members and acquaintances associated with P. Ladefoged, who was later required to match the correct name with each voice sample. These 29 items were presented alongside 13 people whose voices P. Ladefoged had encountered once or twice and 11 voices of individuals that were unknown to him. Each extract varied in duration, with the shortest containing the simple utterance 'hello', to a single sentence and finally 30 seconds of continuous speech. From the 29 familiar voices, 31% were correctly identified from the word 'hello', this increased to 66% from a single sentence to 83% for 30 seconds of continuous speech. The voice recognition tests were only conducted on P.Ladefoged — who himself is a trained phonetician. There are some researchers that claim that voice recognition is better in people who are trained listeners, such as phoneticians (Shirt, 1984 cited in Wilding, Cook and Davis, 2000). Therefore an experiment that looks at the voice recognition performance across a large sample of untrained participants will validate the generalizability of Ladefoged and Ladefoged's (1980) results.

As well as the background characteristics of the listener hampering the generalizability of Ladefoged and Ladefoged's (1980) results, Van Lancker et al (1985) highlight another additional problem with the study. Van Lanker et al (1985) argue that one of the main drawbacks of using personally familiar voices (family members and friends) is that such stimuli tend to produce very few speakers and listeners. A consequence of such a small stimulus set is that memory for familiar voices is likely to be contaminated by strategic processes. With a small set of voices, the participant can use the context provided by the other voices to reject a particular voice. In such cases, it is likely that a voice is recognized because it is *not* one of the other voices (Schmidt-Nielsen and Stern, 1985). Furthermore, there is also a strong possibility of semantic priming within this particular class of stimuli, with the recognition of hearing one's mother's voice facilitating recognition of one's father's voice. As a consequence of semantic priming, the overall recognition rate could be artificially inflated.

To overcome such problems, Van Lanker et al (1985) used a selection of famous voices, such as film stars, etc. This procedure allowed for a larger number of test items to be used, which in turn allowed for a greater number of participants to be tested. In this study, a total of 45 famous voices were used across a range of different voice recognition tasks. Participants were presented with a two second voice sample of each celebrity. For each extract, participants were required to pick out the target name from a choice of six responses. The task was repeated, with the additional change that each voice extract was presented in a backwards format. This particular manipulation was applied to assess whether the acoustic cues that are preserved in backward speech (pitch and pitch range) are sufficient for recognition, even when other cues such as language and temporal structures are distorted.

The results from Van Lancker et al's (1985) study showed that with normal voice presentation rates, participants were successful at correctly picking out the name of the celebrity 69.9% of the time. In the backward presentation condition, performance was reduced by about 12%. This particular voice degradation technique showed that even when language and temporal structures are altered, recognition of a familiar voice can be achieved at around 57.5% of the time.

These findings by Ladefoged and Ladefoged (1980) and Van Lancker et al (1985) show that we can efficiently and rapidly match semantic information to familiar voices. Perhaps one of the main acoustic cues that allow us to carry out this task is pitch and pitch range.

### 1.3: Do faces and voices go together?

Up to now, the studies that have been reviewed have primarily looked at theoretical developments in voice recognition research. This of course, is the main modality of interest for the current purposes of the thesis. Before the person recognition memory models are presented, it would be worth presenting some of the studies which show how faces and voices work together in relation to identity information. This is a particularly interesting point to discuss given the fact that faces and voices are inseparably attached to an individual and as a consequence co-occur extremely reliably (Scweinberger, Herholz and Steif, 1997).

Looking at faces and voices from a developmental perspective, research shows that our ability to recognize others in our visual and auditory environment is present from a very early age. Developmental studies indicate that the cognitive abilities that allow us to recognize voices are available at around 3 days after birth. This ability is of course in stark contrast with our speech perception abilities, which appear considerably later in development (Pinker, 1994). In a study that compared newborns' capacity to recognize voices, DeCasper and Fifer (1980) showed that newborns younger than 3 days of age preferred to listen to their mother's voice than the voice of another female. The baby's performance was measured by recording the length of time the baby sucked on a non-nutritive nipple and comparing the response times for the Mother's and stranger's voice. The overall findings showed that babies would suck for significantly longer periods when played the recording of their Mother's voice thus showing a preference in response to the Mother's voice compared to the stranger's voice. DeCasper and Fifer's (1980) study demonstrates that even after 3 days of postnatal development, a baby possesses the auditory skills that allow them to discriminate between different speakers. Furthermore, a baby can develop a clear preference for their Mother's voice even with limited exposure to their Mother. Technological advancements have shown that the ability to recognize voices seems to be present before birth (Kisilevsky, Hains, Lee, Xie, Huang, Hui, Zhang and Wang, 2003). Kisilevsky et al (2003) showed that foetal heart rate increased in response to the Mother's voice and decreased in response to the stranger's voice.

Although similar observations have been found for facial stimuli (Bushnell, Sai and Mullin, 1989), it appears that the auditory system at the early stages of development is more finely tuned than the visual system. This is reflected by the finding that neonates' abilities to discriminate the face of their Mother from that of a stranger is greatly reduced when peripheral structures of the face, such as the hairline, are concealed (Pascalis, de Schonen, Morton, Deruelle and Fabre-Grenet, 1995). Nevertheless maturational developments in face and voice recognition appear to progress in parallel, with both modalities producing 'developmental dips' in performance at around about 10 years of age before improving and levelling off at around about 14 years of age (Mann, Diamond and Carey, 1979).

There has also been considerable research conducted on the interactive effects of faces and voices. Perhaps one of the most documented studies relating to the interactive effects of faces and voices is the McGurk illusion (McGurk and MacDonald, 1976). The illusion relates to the interference of auditory and visual information in perceiving syllabic sounds. The illusion is created by recombining two separate streams of auditory and visual information. For example, an actor mouths the syllable 'ga', but is dubbed over with the sound 'ba'. The illusion is observed in the blending of the two syllables, so that the viewer perceives the sound 'da'. Interestingly, the intended 'ba' sound is perceived when the viewer does not look at the face. This phenomenon seems to indicate an overriding mechanism in speech perception, such that the visual information overrides the processing of auditory information. Does this processing preference for visual information affect subsequent performance in tests of voice recognition memory?

This research question has been studied by a number of different researchers. For example, McAllister, Dale, Bregman, McCabe and Cotton (1993) used photographs and tape-recorded voices to assess recognition accuracy on subsequent recognition tests. In their first experiment, participants were either allocated to one of two modality conditions. In the auditory-visual condition, participants were given a photograph of the perpetrator to examine whilst they listened to his voice. In the visual only condition, participants were only given the photograph without listening to the tape. Both groups of participants were required to pick out the perpetrator from a visual (sequential) line-up. The initial results showed that there were no significant differences in the visual line-up identification rates as a function of modality. This suggests that hearing a voice does not have an effect on the accuracy of selecting the face at test.

This experiment was modified slightly by replacing the visual only condition with an auditory only condition. In this second experiment, participants were tested with their ability to select the perpetrator from an auditory line-up. The results showed that the identification of the perpetrator was significantly more likely in the auditory-only condition than in the auditory-visual condition.

The findings from McAllister et al's (1993) study showed that visual and auditory information interfere with each other. However, the level of interference is not proportional across the two modalities. Specifically, the presence of a face at encoding is much more likely to have a detrimental effect on voice recognition, than the presence of a voice on face recognition. This asymmetric interference effect has been referred to in the literature as the face overshadowing effect (FOE) (Wilding, Cook and Davis, 2000). It is similar to the verbal overshadowing effect (VOE) observed with faces (Schooler and Engstler-Schooler, 1990). Schooler et al (1990) showed that verbally describing a face after seeing it impairs subsequent recognition of the face (see Macrae and Lewis, 2002 for an alternative account). What differentiates the FOE from the VOE is that the FOE is generated by two simultaneous and involuntary streams of information, whereas the VOE can arise through differential encoding (verbal labels) which is more likely to be under the control of the participant.

An interesting question that stems from the research on the FOE is whether the auditory memory for voices could be improved by the denial of all visual information. There are two possible ways in which this could be investigated. One research route could look at the auditory recognition abilities with people who have visual impairments. The second route could investigate the effects of visually disguising the face during the encoding phase.

Research that has been conducted on the auditory recognition abilities of blind listeners has largely produced mixed results. The rationale behind such research is that people who are blind compensate by showing heightened perceptual powers in non-visual modalities (James, 1890). However, the work by Cobb, Lawrence and Nelson (1979) is not consistent with this claim. They failed to observe a perceptual advantage for blind participants over sighted participants in identifying objects by touch or by sound. On the other hand, Bull, Rathborn and Clifford (1983) showed that blind listeners performed better than sighted participants on an immediate (5 second delay) voice recognition line-up. However, what this particular study by Bull et al

(1983) failed to account for was that within the blind group, degree of blindness and onset of blindness did not relate to voice recognition performance.

Disguising the facial stimulus at encoding has little effect at removing the FOE. As found by Abbott (1999 cited in Wilding et al, 2000), even a stocking disguise, failed to have an impact on removing the FOE. However, it may be the case that this particular disguise still allows for various facial features to be seen and predominate over voice information.

Our ability to recognize others from their facial and vocal cues becomes online very early on in life. Furthermore, evidence from the developmental literature appears to suggest that in the early stages of development our auditory system may be better at recognizing other conspecifics than the visual system. Nevertheless, work from the speech perception domain (McGurk and MacDonald, 1976) as well as the FOE phenomenon (Cook and Wilding, 1997b), suggests that visual cues predominate when visual and auditory stimuli are presented together. This particular asymmetric modality effect has led some researchers to argue that for once-heard voices, the voice recognition route in person memory models may not work as efficiently if a face has been presented at first voice exposure (Cook and Wilding, 1997b). A further finding that needs to be accounted for by these person recognition models is to what extent are face and voice recognition routes independent? This question is raised by the robustness of the FOE, which suggests that these two separate streams of information must be combined at some point in the person recognition memory system.

### 1.4: The voice area.

The aim of this section is to summarize the neuropsychological and neuroimaging data relating to the theoretical developments of the voice-selective areas of the human brain.

Bodamer (1947 cited in Neuner and Schweinberger, 2000) is often cited in the psychological literature as the first researcher to introduce *prosopagnosia* as a separate form of agnosia confined to faces. Patients diagnosed with the disorder show a profound inability to recognize other individuals from their face. However, introspective reports from patients suggest that most prosopagnosic patients deal with this deficiency by recognizing people from their voice or from other non-facial characteristics (Sacks, 1986). More centrally relevant to this thesis are the neurological case studies that report a similar form of agnosia in the auditory domain. In such cases, patients are able to recognize and name auditory sounds, however they show a marked inability to recognize other people from their voice. This form of agnosia is referred to by Van Lancker, Cummings, Kreiman and Dobkin (1988) as *phonagnosia.*

The search for voice selective areas of the human brain are often traced back to Van Lancker and Kreiman's (1987) claim that the right and left hemispheres are differentially involved in the recognition of familiar and unfamiliar voices, respectively. This claim by Van Lancker and Kreiman (1987) was made in light of their argument that voice recognition (recognizing friends, relatives and famous voices) is a Gestalt process governed by properties of the right hemisphere. Whereas voice discrimination (unfamiliar voices) is more likely to be associated with a featural process related to left hemisphere functioning. To gather data for their Gestalt vs. featural processing account, Van Lancker and Kreiman (1987) studied the performance of patients with different localized lesions on tests of voice recognition (famous men) and voice discrimination (unfamiliar voices).

In line with Van Lanker and Kreiman's (1987) arguments, patients with right hemisphere brain damage (RBD) did poorly on the voice recognition task. This observation is consistent with the view that the pattern recognition properties of the right hemisphere facilitate voice recognition. Both left hemisphere brain damage (LBD) patients and (RBD) patients showed impairments on the voice discrimination task, thus hinting that unfamiliar voice discrimination may not purely rely on featural processing, but on an interaction of both a featural (left hemisphere) and a pattern recognition (right hemisphere) processes. Furthermore, LBD individuals did not differ from controls on the voice recognition task, again highlighting the importance of an intact right hemisphere for this particular voice ability.

This attempt to localize the different voice processing abilities to different sides of the brain is undoubtedly a significant step forward to elucidating the voice selective areas of the brain. However, in the context of modern advances in technology, the view that voice recognition and discrimination can be reduced to right vs. left hemisphere processing is rather over simplistic. Similarly, the Gestalt vs. featural characteristics of the right and left hemispheres has begun to lose conceptual appeal (Boles, 1984).

Perhaps one of the most extensive reviews of person recognition memory impairments was produced fairly recently by Neuner and Schweinberger (2000). The remarkable aspect of Neuner and Schweinberger's (2000) review is that it contains a simultaneous assessment of face, voice and name recognition abilities in the same group of patients (N = 36). Along with tests of face, voice and name recognition, patients were tested with corresponding object recognition tests (pictures of objects, environmental sounds and common words). For purposes of relevance, Neuner and Schweinberger's (2000) data for patients who only exhibited voice recognition problems will be summarized.

There were two voice recognition tests given to patients. One involved 64 voice samples, half of which belonged to famous individuals in the German media and half of which contained non-famous voices. Each sample was 2 seconds long in duration. For each voice, participants were required to state if they found it familiar or unfamiliar. If they found it familiar, they were asked to produce the person's name. The second task involved a voice-matching test. This test included a selection of 54 unfamiliar voices. The test involved listening to two separate samples and deciding whether they belonged to the same or different speaker. On half the trials the speakers were the same, on the other half of the trials the speakers were different.

From the 36 patients that were assessed, 10 patients showed significant impairments in at least one form of the person recognition test that was not accompanied by impairments in the corresponding object recognition tests. From this selection, a sample of 4 patients showed impairments that were confined to voices. This deficit occurred even though the patients showed a preserved ability in face and name recognition. Furthermore, their deficit can not be attributed to a deficit in general auditory processing, since all 4 patients' performance was well within the range of the control group for sound naming and recognition.

Three of the four patients performed well within the range on the voice-matching test. The remaining patient exhibited poor overall performance in the voice recognition test as well as the voice-matching test. The pattern of performance shown by this patient may indicate that his difficulty lies in the perceptual encoding of voice specific cues. In the four patients that demonstrated selective impairments in voice recognition, all but one suffered from a lesion to the right hemisphere. The right hemisphere damage in these patients is reminiscent of Van Lancker and Kreiman's (1987) framework. However, a more detailed explanation is needed to understand the neurological underpinnings of voice discrimination. A simple left hemisphere / featural explanation can not account for Neuner and Schweinberger's patient (case 1) who showed a deficit in matching unfamiliar voices, yet the localization of his lesion was to the right. Similarly case 24 did not show a deficit in unfamiliar voice matching despite contracting a lesion to the left side of the brain.

Neuropsychological case studies are inarguably a valuable source of evidence in unravelling the neural correlates of voice recognition and discrimination in the human brain. A complimentary line of converging evidence comes in the form neuroimaging studies which have allowed researchers to assess pre-morbid neural correlates in cognitive assessments of voice processing.

For example, Shah, Marshall, Zafiris, Schwab, Zilles, Markowitsch and Fink (2001) used functional magnetic resonance imaging (fMRI) to localize brain regions associated with the processing of familiar faces and familiar voices. The stimuli that were presented to participants included familiar and unfamiliar faces and voices. The familiar items were photographs and recordings of relatives and friends of each individual who took part in the experiment.

The imaging data revealed that the neural activity in response to seeing faces (irrespective of familiarity) was localized on the fusiform gyrus bilaterally — a finding consistent with the 'fusiform face area' (Kanwisher, McDermott and Chun, 1997). Neural activity in response to voices was associated with the superior temporal gyrus bilaterally, extending into the primary and secondary auditory cortex bilaterally. Activation in response to personal familiarity (irrespective of stimulus modality) was observed bilaterally in the posterior cingulated cortex including the retrosplenial cortex. The authors proposed that this multimodal region of activation could be associated with familiarity assessments and possibly in identifying people.

In another fMRI study, Belin, Zatorre, Lafaille, Ahad and Pike (2000) aimed to compare the neural responses to vocal sounds (vowels, words, laughs and coughs) with non-vocal sounds (natural sounds and animal cries) in a sample of healthy participants.

The analyses showed that vocal sounds generated significantly greater activation than non-vocal sounds bilaterally in several regions of the non-primary auditory cortex. Voice-sensitive activation was found along the upper bank of the central part of the superior temporal sulcus (STS). The STS is a deep, long sulcus that runs along the whole temporal lobe and is also found in many non-human primates. Furthermore, the level of voice-sensitive activity appeared stronger in the right hemisphere and was distributed in three bilateral clusters along the STS.

In their second experiment, Belin et al (2000) compared STS activity for speech and non-speech stimuli with a different set of auditory control stimuli. These included bell chimes and other human non-vocal sounds such as finger-clicks and handclaps. Activation of the STS was significantly greater for the vocal stimuli (speech and non-speech) than for the control stimuli combined. This pattern of results illustrates a number of interesting findings. Firstly, it shows that the voice-selective activation of the STS is not simply activated by the presentation of exemplars of a single category (i.e.; bells). Secondly, the STS does not merely respond to any sound of human origin — thus highlighting the voice selective nature of the STS.

Furthermore, the level of STS activity appears to be modulated by frequency properties of the auditory stimulus. The response of the STS to vocal stimuli is decreased by filtering out either the high or low frequency components of the vocal stimulus. The decrease in signal activity as a function of frequency removal is consistent with Belin et al's (2000) behavioural data. Specifically, participants' accuracy on a vocal / non-vocal decision task and gender identification task decreased as a function of frequency removal.

The voice selective nature of the STS in the human brain has numerous implications for evolutionary and comparative research. The study illustrates that whilst language may be a unique human quality, the ability to extract identity related cues from vocalizations is a quality humans share with other species (see Insley, 2000).

The review of the studies in this section has traced the theoretical and empirical developments of the neurological correlates of voice recognition. The review showed that whilst voice recognition and discrimination may be two cognitively different tasks, the localization of these tasks can not be simply reduced to right vs. left hemisphere processing (Van Lancker and Kreiman, 1987). Neuroimaging techniques suggest instead that the neural correlates involved in voice recognition and discrimination may involve several intricate cortical areas. Whilst these areas are voice specific, there is some evidence to suggest that this voice selective node feeds into a multimodal centre which may be responsible for familiarity assessments and person identification (Shah et al, 2001).

### 1.5: Voice as evidence.

Voice evidence is already being used in English criminal trials as a legitimate source of evidence to establish identity. As noted by Ormerod (2001), voice evidence can form a part of almost any criminal case. The guidelines that are set in place to assess such evidence largely follow the rules that govern eyewitness procedures.

Voice evidence can be differentiated in terms of evidence used as recognition vs. identification of the suspect. This point is similar to Van Lancker et al's (1987) recognition / discrimination distinction. Recognition refers to those instances in which the witness has prior familiarity with the voice of the offender, whereas identification refers to those instances in which the voice was not previously known to the witness. There have been documented cases that have used voice identification evidence (R. v R, 2000).

Other cases that have admitted evidence on the basis of a single auditory exposure of the perpetrator include State v. Hall (1928 cited in Ormerod, 2001). The victim was attacked by a perpetrator, who attempted to conceal his identity by wearing a hooded disguise. The victim was able to identify the perpetrator's voice as belonging to a customer at his pool room who he had encountered a few hours before on the evening of the attack.

Voice evidence is particularly beneficial when the nature of the crime occurs purely through non-visual means such as bomb threats, ransom demands and other threatening behaviour made through the use of a telephone. In the kidnapping case R v. Robb (1991), a wealthy business man was kidnapped and ransom demands for his safe return were made by telephone to his wife. The telephone calls were subsequently recorded. A closer inspection of the taped recordings revealed that the voice making the ransom demands belonged to the victim, who was later charged with conspiracy to commit false imprisonment.

At the time of writing this thesis, an investigation is still underway to identify the murderer of Lieutenant Colonel Workman who was shot on the doorstep of his home. The only lead police officers have to work on is a 60 second phone call made to the emergency services by an anonymous caller. The caller has not come forward and the village's residents have been provided with a recording of the voice to see if they can recognize it as belonging to a local resident.

To date, the development of voice as evidence guidelines for the English courts is still in progress. The admissibility of such evidence varies largely from country to country. For example, even a little over a decade ago, a U.S judge asserted that there was no scientific evidence that voice identification was as dangerous as visual identification (State v. Burnison, 1990 cited in Ormerod, 2001). In contrast, Indian law permits the admission of recognition but not identification evidence (Kirpal Singh v. State of Uttar Pradesh, 1963 cited Ormerod, 2001).

To conclude, the cases presented in this section demonstrate how the human voice is readily accepted by some courts as a form of identity and illustrate the need to further our understanding of our voice recognition / identification abilities. In cases that rely on eyewitness evidence, legal proceedings are often accompanied by the Turnbull warning given to the jury by the judge. This warning brings to the attention that eyewitnesses giving evidence may well be mistaken. Ormerod (2001) argues in favour of developing a similar style warning for voice evidence. However, the exact phrasing of this warning needs to encapsulate a wide range of factors that affect person recognition / identification evidence and must also capture the factors that are voice specific. For example, unlike face identification, the effects of stress of violent crime greatly affects the speaker's voice (Saslove and Yarmey, 1980) which in turn will differ how the voice is encoded and subsequently detected in a voice line-up parade.

## 1.6: Person recognition memory models —boxes and arrows.

Throughout this thesis, the term person recognition memory shall be used in relation to people's abilities to recognize other human beings from their face and / or voice. Whilst the Bruce and Young (1986) (see figure 1.6a) model of person recognition does implicitly incorporate a voice recognition route, much of the work that has been accumulated over the years has placed greater emphasis on the facial recognition route. The most recent attempt to accommodate a voice recognition route in Bruce and Young's model (1986) has been proposed by Belin et al (2004). Their proposal is a direct mapping of the face recognition route for voices (see figure 1.6b). Research from the voice recognition domain needs to be carefully evaluated in order to see whether the voice recognition route can be conceptualized as a straightforward analogy of the face recognition route.

Bruce and Young (1986) developed a functional model of face recognition in light of the extensive work that had already accumulated in the face recognition literature. This is not to say that the box and arrow model proposed by Bruce and Young is the first of its kind to integrate this work into a theoretical framework. At around about the same time, a number of different researchers were attempting to integrate the face recognition research into a comprehensive cognitive model (Damasio, Damasio and Van Hoesen, 1982). However, there is general agreement in the literature that the model outlined by Bruce and Young is prototypical (Rapcsak, Polster, Comer and Rubens, 1994). Bruce and Young's model is displayed in figure 1.6a. Figure 1.6b, shows a simplified version of Bruce and Young's model with a proposed route for voices (Belin et al, 2004). As can be seen from figure 1.6a, there are a number of different boxes representing the main components involved in face processing.

Figure 1.6a: Bruce and Young's (1986) Functional Model of Face Recognition.

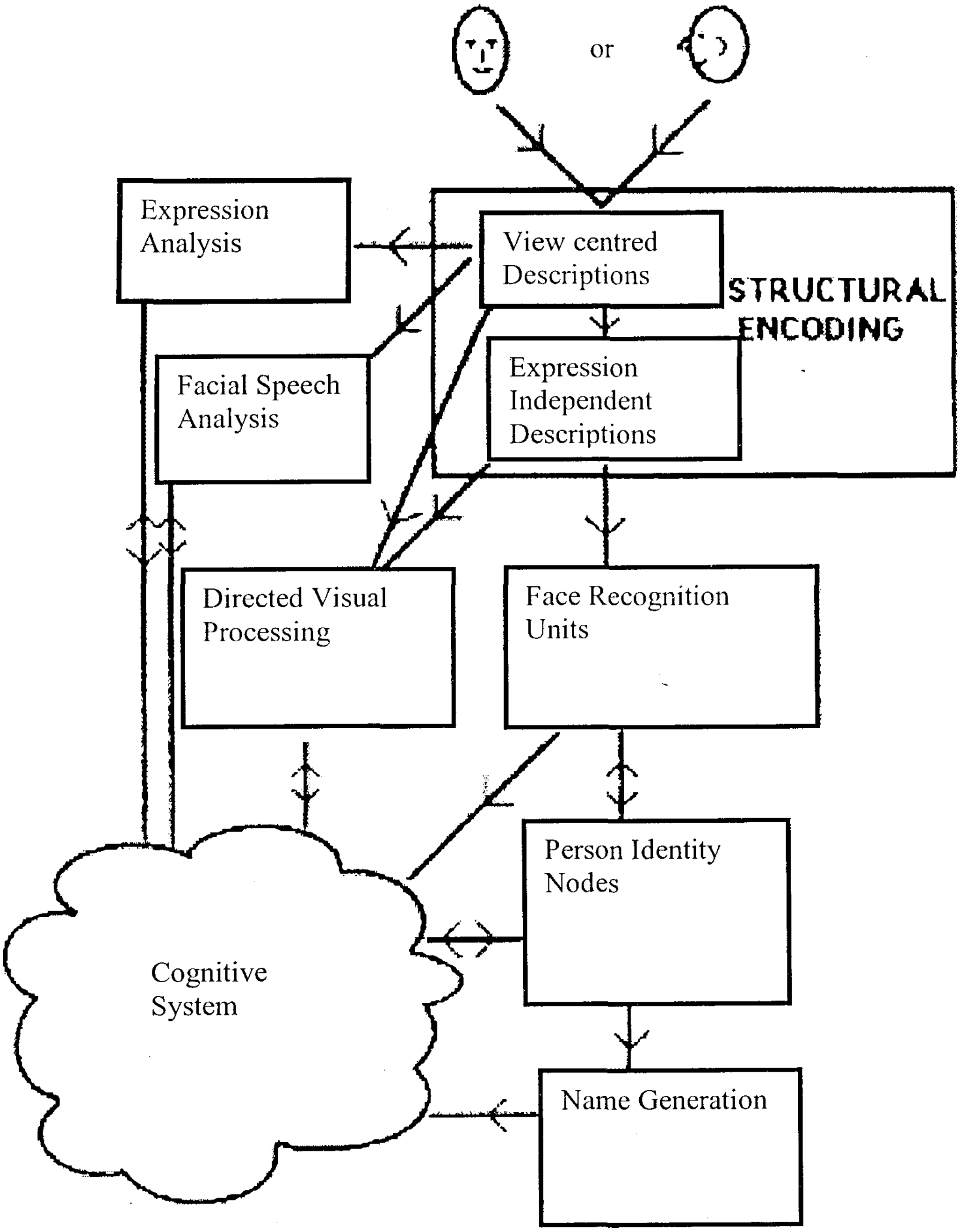
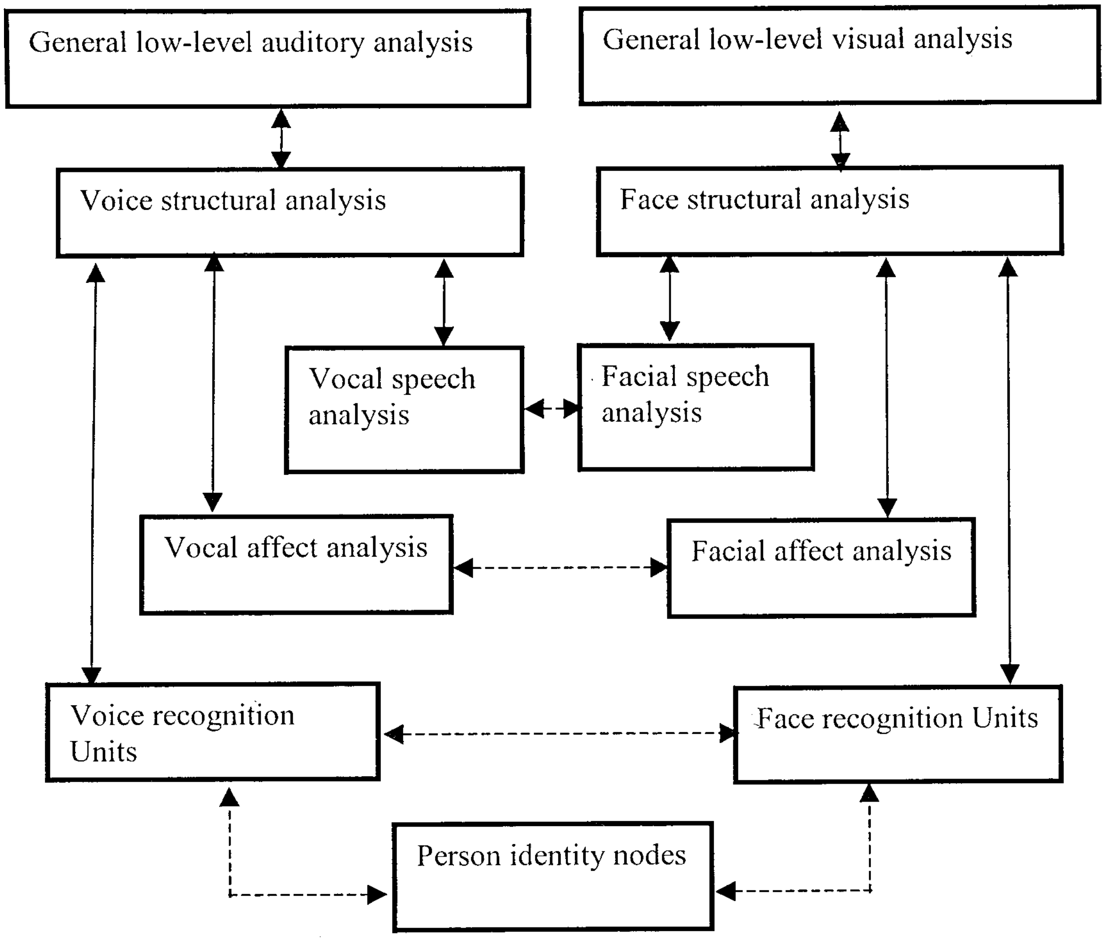


Figure 1.6b. Simplified version of Bruce and Young's (1986) face recognition route (right) with proposed voice recognition route by Belin, Fecteau and Bédard (2004) on the left. Dashed arrows indicate multimodal interactions.



This box and arrow schematic follows on from similar models from object (Marr, 1982) and word (Morton, 1979) recognition. The general consensus shared by such models is that a 'box' represents any processing module that plays a distinct functional role. Specifically, the 'box' or module in question, can be isolated through various experimental manipulations or as a result of brain damage. The arrows represent access of information and the activation of one component by another. Following convention, the cognitive system in this model is represented in the form of a cloud. The functions of the cognitive system are described in a separate section in this chapter. In the proceeding sections, the model will be described in terms of the perceptual and cognitive aspects of face recognition.

From a perceptual stance, the model is concerned with the processes involved in expression and facial speech analysis. Whilst these components play an important role in conveying what someone is feeling and what they are saying, these components in themselves do not form part of our central social repertoire. It is the cognitive aspects of the model, that is the ability to retrieve personal information about a known individual, that forms the central impetus of the model and this thesis in particular.

According to Bruce and Young (1986), our daily face recognition skills are governed by three stages. These stages involve familiarity, semantic information and name retrieval. These are all functionally distinct components that need to operate in a sequential manner in order for face recognition to be successfully achieved. Although the authors acknowledge that person recognition can be achieved through other social cues such as the person's voice or gait, the emphasis of their paper is on visual face recognition. The following sections will outline the three main stages involved and summarize some of the main studies that have been used to support the various components and the overall sequential structure of the model.

The first stage in the Bruce and Young model involves structural encoding. For the recognition of familiar faces, this is seen to take place at the level of the Face Recognition Units (FRU). There is one FRU for each known person and they contain abstract descriptions or codes for each face. These abstract descriptions include descriptions of local and / or distinctive features of the face and its configuration. It is proposed by the authors that there are corresponding voice recognition units (VRUs) that are activated when recognition is based on vocal cues. Returning to the role of the FRUs, the level of activity of each given FRU is relayed to the cognitive system which compares the strength of the signal between the stored description of the person's face and the facial input received. The strength of this similarity index is used as the basis for deciding whether or not the face is familiar.

Familiarity in itself, is not enough to tell us definitively to whom the face belongs to. This can only be achieved through the retrieval of an identity-specific semantic code. These types of codes include descriptions of the person's occupation, where they are usually encountered, who they are associated with and so on. Bruce and Young (1986) contrasted these identity-specific semantic codes with visually derived semantic codes (see acoustically derived semantic codes, page 4). Visually derived semantic codes can be created for unfamiliar and familiar faces. They are usually generated by physical properties of the face, such as sex, age and expression. Non-physical judgments can also be formed, for example judging the honesty and intelligence pertaining to a particular face. Whilst the visually derived semantic codes enable us to understand and judge the person's age and sex, it is the retrieval of the identity-specific semantic codes that allows us to achieve a sense of knowing to whom the face belongs to.

According to Bruce and Young (1986), these identity-specific semantic codes can be accessed by the relevant FRU. The identity-specific semantic codes are held in a portion of associative memory which Bruce and Young (1986) have labelled as person identity nodes (PINs) in their model. There is one PIN for each person and this stores the identity-specific semantic codes and this allows us to achieve a sense of *person* recognition, as opposed to face recognition. It is through this route that semantic knowledge is activated. Additionally, the PINs can also be accessed by other person-specific cues, such as the voice or name.

The final and perhaps most complete stage in face recognition is the retrieval of the person's name. The Name Generation box represents the lexical output system of the model. This contains the name code that allows us to retrieve the name for the face. Crucially, this final stage of the face recognition process can only be accessed through the PINs.

Evidence in support of Bruce and Young's (1986) box and arrow model is impressive. Latency data, for example, has provided evidence for the sequential structure of the model. These laboratory based studies have shown that the feeling of familiarity occurs before an occupation is retrieved (Young, McWeeny, Hay and Ellis, 1986). Moreover, the retrieval of an occupation to a face is faster than the retrieval of a name to a face (Young, McWeeny, Ellis and Hay, 1986). Binary matching tasks offer additional support for the sequential nature of the model. In such studies, the participant's reaction time in deciding whether two items share the same property is measured. For example, Young, Ellis and Flude (1988) found that deciding whether two faces shared the same occupation was made significantly faster (745ms) than deciding whether the two faces shared the same name (803ms). These reaction time studies are consistent with the structure of the model as familiarity will be the first stage of the model which is activated, thus giving faster response times. In the case of naming, as this is in the last stage of the model, reaction times should be the slowest for this type of category response compared to occupations, which occurs prior to naming (see figure 1.6a).

Additional laboratory work has looked at participants' abilities to learn associations between unfamiliar faces. McWeeny, Young, Hay and Ellis (1987) showed that participants were significantly better at learning the face-occupation associations to unfamiliar faces than they were in learning the face-name associations. The precedence for occupational associations continued even when the occupations and names were nominally identical. This latter observation is referred to in the literature as the 'Baker-baker' paradox by Cohen (1990b).

Like reaction time studies, error patterns reveal an ordering among the stages of processing. In a study by Young, Hay and Ellis (1985a), twenty-two participants were each asked to keep a diary and record any problems they had encountered in recognizing people over an eight week period. In total, there were 922 errors recorded, with the lowest scoring participant reporting 13 incidents.

Two of the most common types of person recognition failure were classified by the authors as (i) failure to 'place' the person and (ii) failure to recall full details about the person. With regards to the first type of error, there were a total of 233 incidents recorded. This type of failure is characterized by a strong feeling of familiarity elicited by the face and a strong sense of certainty that the face has been encountered before — however no other information can be associated with the face. Specifically, participants have problems recalling when or where they had previously encountered the person before. This type of error is also known as the 'familiar-only' error.

The second type of error reported in Young et al's study (1985a) produced 190 incidents. In such cases, the face was associated with some contextual information, however the person's name was temporarily inaccessible. This type of error is also most commonly referred to in the literature as the tip-of-the-tongue (TOT) state. It has been addressed experimentally with real and low frequency words by Brown and McNeill (1966). What makes this error particularly fascinating is that it is not necessarily an all or nothing process. As Brown and McNeill (1966) noted, despite being unable to name the word from its definition, participants were able to retrieve partial information about the characteristics of the word such as the number of syllables, the initial letter or sound and the stress pattern. The generation of partial information has also been observed with real names (Yarmey, 1973).

Both types of errors can be readily accommodated into Bruce and Young's (1986) model. The failure in placing the person is interpreted in terms of activation of the relevant FRU — which elicits a feeling of familiarity. However in this particular case the FRU has failed to activate the relevant PIN, which contains the relevant information about where the person is usually encountered and what they do, for example. The TOT state occurs when the relevant PIN has been successfully activated and the semantic information that it contains has been retrieved. However, access to the name code, which is stored in a separate component of the model is blocked.

Whilst the diary study by Young, Hay and Ellis (1985a) provides a naturalistic way of observing the types of problems we encounter in recognizing faces, similar patterns of behaviour have also been observed in the laboratory in the form of cueing experiments (Hanley and Cowell, 1988). A cueing experiment in person recognition memory is developed around the premise that certain cues will help to resolve the specific types of person recognition failures by allowing access to the relevant component of the face recognition model.

In their cueing experiment, Hanley and Cowell (1988) showed that participants who had difficulty in placing the person found biographical cues most efficient in helping them to identify the person, since this cue helped them move onto the next stage of the model, the PINs. Participants in tip-of-the-tongue states found the name initial most beneficial, as this helped them move from the PIN level to the name retrieval stage. Similar patterns of performance have also been observed in the recall of names of famous voices (Schweinberger, Herholtz and Sommer, 1997).

Data from these cueing experiments (see also Brennen, Baguley, Bright and Bruce, 1990) adds particular merit in terms of the predictive value of the model, which in turn helps to reinforce the main structural and functional characteristics of the model. Collectively, the data suggest that names are stored separately from other information that we know about a person, such as their occupation. This means that the retrieval of someone's name is conditional upon the retrieval of biographical information about the person. In other words, you have to go through the PINs in order to gain access to the name code. This hierarchical structure is sometimes referred to in the literature as the semantics-name sequence or SNS (Brennen, David, Fluchaire and Pellat, 1996).

Interestingly, Bruce and Young (1986) also highlight the types of person recognition errors that should not occur. In particular, they argue that there should not be any instances in which a face is named in the absence of any identity specific information. Structurally, this is forbidden by the model as there is not a direct route from the structural codes to the name codes in the name generation component. At the time of developing the model, Bruce and Young (1986) cited the patient of Williams and Smith (1954), who managed to name personally familiar faces (fellow military officers) without being able to specify when and where he had met them. Bruce and Young (1986) did not go into great lengths in explaining this performance by Williams and Smith's (1954) patient, as he was also diagnosed as amnesic — thus making this type of error demonstrated by this patient difficult to interpret.

Returning to the diary study by Young et al (1985a), even in the corpus of responses collected by these researchers there were no accounts in which a participant could name someone in the absence of identity-specific information. Subsequent diary studies have also failed to observe these contentious types of errors (Schweich, van der Linden, Br dart, Bruyer, Nelles and Schils, 1992). However, this 'naming without semantics' phenomenon has recently resurfaced in the neuropsychological assessment of Mme.DT (Brennen, David, Fluchaire and Pellat, 1996). This patient was able to recall the names of some famous French celebrities even though she could not retrieve their occupational descriptions (but see Hodges and Greene, 1998, for a critique).

Work from the neuropsychological field has also provided an important source of data in support of Bruce and Young's model (1986). Evidence to support the view that the route involved in the processing of familiar faces is distinctly different from the route involved with unfamiliar face processing has been observed in patients with prosopagnosia. The prosopagnosic syndrome is often defined as the inability to recognize familiar persons by their face alone (Bodamer, 1947 cited Neuner and Schweinberger, 2000). The disorder can be classified further as cases that are apperceptive and associative. The apperceptive form of prosopagnosia is considered to be a perceptual deficit, in that the patient demonstrates an inability to encode facial traits. The associative form of prosopagnisia is associated with intact perceptual facial encoding coupled with an inability to access information about a person from their face. Malone, Morris, Kay and Levin (1982) described two patients, the first of which remained impaired in the visual analysis of unfamiliar faces whilst their ability to recognize familiar faces improved. The second patient, on the other hand, showed a marked improvement in the processing of unfamiliar faces in the presence of their prosopagnosic deficit. The pattern of performance observed by these patients demonstrates two functionally distinct routes in processing familiar versus unfamiliar faces.

There are also neuropsychological studies that look at the post-structural difficulties that some patients display. Flude, Ellis and Kay (1989) outlined the case of EST who displayed a selective deficit in naming familiar faces. This was observed even though EST performed normally in tasks testing the structural encoding of faces and the retrieval of identity-specific semantic information. The pattern of performance exhibited by EST is therefore consistent with Bruce and Young's notion that names are stored separately from other semantic information about a person.

To conclude, the converging evidence that has amassed over the years has helped to reinforce the various functional aspects of the box and arrow model as well as to support its hierarchical architecture. This in turn has not only helped to elucidate the various stages involved in normal face processing, but also contribute to a better understanding of what is lost and / or spared after brain trauma. The closing paragraphs will address the additional 'cloudy' component of the model — the cognitive system and its role in face processing.

## 1.7: The Cognitive System.

Perhaps the most under specified component of the Bruce and Young (1986) model is the cognitive system, which the authors argue is on a par with other 'catch-all' components in cognitive models such as the central executive (Baddeley, 1997). The main duty of the cognitive system is to carry out a supervisory role by being involved in the planning and decision-making of the various components within the model.

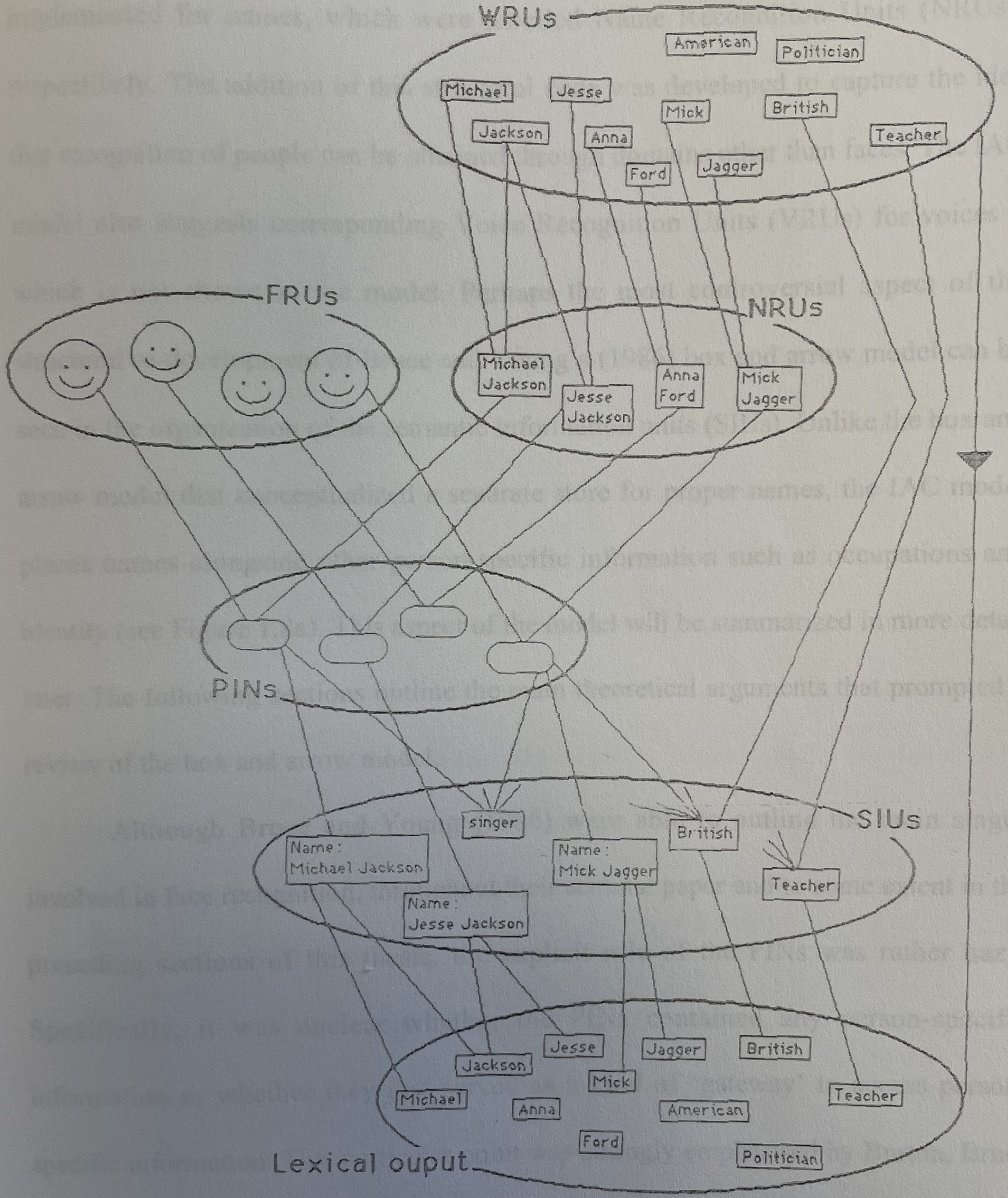
One specific role that has been attributed to the cognitive system is the 'familiarity check' process as outlined by Ellis (1983). Such a 'familiarity check' process can be used to resolve 'resemblance' experiences. This occurs when we notice that a face holds a very strong resemblance to somebody else. Even though the FRU is activated, it is the role of the cognitive system to set up a decision criterion to accept or reject whether the activity from the FRU is enough to deem the person as somebody we are familiar with. Damage to this 'familiarity check' process of the cognitive system is believed to underlie some misidentification errors, as observed in patient M.R (Ward, Parkin, Powell, Squires, Townshend and Bradley, 1999).

Bruce and Young (1986) also view the cognitive system to be the storehouse of episodic memory. The information stored within episodic memory differs from the information stored in the PINs. The PINs contain uniquely identifying semantic knowledge of people, such as 'won an Oscar for Gladiator' (Russell Crowe). Episodic memory, on the other hand tends to store more general and episodic based information about the person such as details of films that they starred in. Access of the PINs therefore is more likely to facilitate the identification of Russell Crowe, than merely knowing that you may have watched him at the cinema.

## 1.8: Burton, Bruce and Johnston — an interactive activation and competition (IAC) model.

The box and arrow model by Bruce and Young (1986) viewed successful face recognition to be achieved through the successive access of the Face Recognition Units (FRUs), Person Identity Nodes (PINs) and Name retrieval. The (IAC) model extends these core structures within a connectionist architecture (McClelland and Rumelhart, 1981). The model is now represented as a pool of simple processing units which contain inhibitory units. Each pool is connected to each other via excitatory links. All links between the components are of equal strength and are bi-directional, which allows activation to pass within the model in a cascade fashion (Figure 1.8a). This means that processing at one stage of the model need not be completed before the next stage can be initiated. This of course, is in stark contrast with the strict sequential nature of the box and arrow model.

Figure 1.8a. The IAC model of face recognition taken from Burton and Bruce (1993).



At a surface level, the IAC model holds many similarities to the box and arrow model. The FRUs are retained in the IAC model — activation at this level of the model corresponds to the classification of the face. Similar recognition units are also implemented for names, which were labelled Name Recognition Units (NRUs), respectively. The addition of this structural code was developed to capture the idea that recognition of people can be obtained through domains other than faces. The IAC model also suggests corresponding Voice Recognition Units (VRUs) for voices — which is not shown in the model. Perhaps the most controversial aspect of the structural re-development of Bruce and Young's (1986) box and arrow model can be seen in the organization of the semantic information units (SIUs). Unlike the box and arrow model that conceptualized a separate store for proper names, the IAC model places names alongside other person-specific information such as occupations and identity (see Figure 1.8a). This aspect of the model will be summarized in more detail later. The following sections outline the main theoretical arguments that prompted a review of the box and arrow model.

Although Bruce and Young (1986) were able to outline the main stages involved in face recognition, throughout their seminal paper and to some extent in the preceding sections of this thesis, the explicit role of the PINs was rather hazy. Specifically, it was unclear whether the PINs contained any person-specific information or whether they just served as a kind of 'gateway' to access person-specific information. This particular point was strongly emphasised by Burton, Bruce and Johnston (1990) who argued that 'Bruce and Young did not explicitly separate semantic information from the PINs, since they did not spell out whether semantic information about person identity is stored 'in' the PINs or accessed 'via' the PINs' (1990, p. 365).

The first advantage of the IAC model is its specification of the PINs. Here, the authors specifically state that the PINs are nodes that allow access to semantic information about a person which is stored in the Semantic Information pool. The PINs are content-free: their role is to gather information from the modality-specific units (FRUs, NRUs and VRUs) and provide access to the semantic information units in semantic memory. Unlike Bruce and Young (1986), the PINs were not viewed to contain stored knowledge about a person. The new role of the PINs is now seen to be the locus of familiarity decisions, unlike the Bruce and Young model which envisaged familiarity decisions to lie at the level of the structural codes. Moving the locus of familiarity from the structural codes to the PINs accommodates the more parsimonious observation that we encounter people in more than one modality. Familiarity is achieved when any PIN reaches a common threshold. This means that the same mechanism is used to make familiarity decisions for all person familiarity judgments — faces, names and voices. Since the PINs act as the converging point for all modalities, information available in one modality (faces) is also available in all modalities (voices). In other words, semantic information is not modality specific.

The familiarity-only error described in previous sections is captured in the IAC model as a block between the PINs and the associated links in the Semantic Information Units. This means that the threshold for familiarity in the PINs is reached, triggering a sense of familiarity, however the level of activation passed on to the SIUs is not sufficient to retrieve the identity-specific information. This particular explanation is consistent with the pattern of performance observed by M.E (de Haan, Young and Newcombe, 1991). M.E was able to successfully rate the familiarity of faces and names at a comparable level to the control group. However, M.E demonstrated problems in identifying the names and faces by means of semantic information. Interestingly, of the few faces that she could associate some semantic information to, she was able to provide the correct name. This suggests that M.E does not have a name retrieval problem, her specific problem lies therefore in the retrieval of identity-specific semantic information. When she was asked to match a name to a face that she could not previously provide semantic information to, she managed to score 23 out of 26 on this face-matching task. M.E's performance therefore reflects a heightened form of the familiarity-only error reported in the normal population (Young, Hay and Ellis, 1985).

The earliest stage in the Bruce and Young (1986) model at which face-name matching could be achieved is at the PINs level — thus suggesting that M.E has intact PINs. Intact structural codes alone would not allow her to succeed on this task. However, this interpretation alone can not be the whole story given the way in which the familiarity-only error is represented in Bruce and Young (1986) model. According to Bruce and Young (1986) this type of error occurs because the relevant FRU has fired (signalling familiarity) but access to the PINs in temporarily inaccessible. This interpretation is particularly problematic because if M.E was using the PINs to perform within the normal range on the matching task, then she should have also been able to recall semantic information about the people. This inconsistency highlights the importance of separating the PINs from the Semantic Information Units pool — a conceptualization which was largely neglected in the Bruce and Young (1986) model. M.E's inability to identify people occurred with faces *and* names. Interestingly, she was able to correctly make familiarity judgments to these items on preceding occasions. This pattern of performance is consistent with the IAC model's view that familiarity is determined at the level of the PINs. Therefore, M.E's deficit reflects an access problem from the PINs to the SIUs. This is a pattern of performance that can be elegantly captured by the IAC model.

A final difference between the Bruce and Young model (1986) and the IAC model relates to name retrieval (Burton and Bruce, 1993). As can be seen in figure 1.8.a, names are stored alongside other semantic information about the person rather than in a separate store as originally conceptualized by Bruce and Young (1986). Under the IAC model, names are difficult to retrieve because they are unique (Burton and Bruce, 1990). Occupations and other biographical features (nationality) can be shared by many individuals (labour Prime Minister and British), yet we only know of one Tony Blair. The difficulty in retrieving names is therefore a direct result of the architecture of the model. Specifically, names because of their 'uniqueness' generally have one link compared to several for other biographical features. The bi-directional nature of the links means that activation will be the greatest for the units with most links. As a result, there will be more activation and better chance of retrieval for units with multiple links than single links.

Although such a notion of name retrieval has been readily simulated (Burton and Bruce, 1992) and can elegantly explain developmental differences in name retrieval (Scanlan and Johnston, 1997), other researchers have opposed this idea. These researchers have argued that name retrieval units should be stored in a separate pool either in parallel with SIUs or following them (Hanley, 1995).

## 1.9: Aims of thesis.

The review of the existent literature from sections 1.1 to 1.8, has shown that whilst person recognition memory models include an auditory route for person recognition, very little attempt has been made to explicitly integrate the research done on voice recognition (sections 1.1 — 1.5) within these models. This is most apparent in sections 1.6 to 1.8, which illustrates the predominant use of face recognition data to evaluate person recognition memory models. Furthermore, other researchers (Wilding et al, 2000) argue that simply adding an auditory route in these models is not an adequate analogy to the face recognition process. Specifically, they argue that the face recognition route may have preferential links to semantic information whereas the voice recognition route may predominantly be used to aid communication rather than identification.

There have been some attempts to assess whether the voice recognition route in these models works in parallel to the face route (Schweinberger, Herholz and Sommer, 1997). However, there have only been two studies that have explicitly looked at the recognition of familiar faces and voices within the same experiment (Hanley, Smith and Hadfield, 1998 Hanley and Turner, 2000). The purpose of this thesis is to extend the work done by Hanley, Smith and Hadfield (1998) by firstly replicating their findings of a cross-modal difference in person identification and secondly by assessing whether this difference can be removed in terms of employing a different experimental design. The specific feature of this thesis is that face and voice recognition will always be compared in the same series of experiments. The empirical overview of this thesis is as follows:

**Chapter 3:** This chapter will investigate how participants make familiarity judgments to faces and voices and whether there are differences between the two modalities in terms of how efficiently they elicit person specific information. The specific area of interest is the familiar only error (Young, Hay and Ellis, 1985a). In doing so, the chapter will review the role of the PINs in person recognition judgements and include a summary of the findings from face and voice priming experiments. The critical point of investigation in this chapter is whether the change in experimental design can remove any cross-modal differences in person specific identification. This will be done by introducing the concept of the 'mixing effect' from the confidence-accuracy literature (Hollins and Perfect, 1997).

**Chapter 4:** This chapter will replace judgments of familiarity with judgments of confidence in order to assess whether there are any cross-modal differences between faces and voices when participants are required to use confidence judgments. The review of the studies will focus on the findings from the confidence-accuracy literature (C-A) in person recognition memory. The use of confidence judgements will also serve to generate Receiver Operating Characteristics (ROCs). The chapter will review the study of the ROCs procedure and outline how they can be used to infer differences in recognition processes (Yonelinas, 1999).

**Chapter 5:** This chapter will aim to discover whether face recognition can be reduced through visually degrading the facial image. The chapter will highlight the established techniques in the face recognition literature and evaluate whether these techniques can produce overall levels of performance in the face condition that are similar to the voice condition. Furthermore, the chapter will extend the use of the ROCs procedure to assess whether changes in recognition processes are altered by changes in accuracy.

**Chapter 6:** This chapter will assess whether cross-modal differences can be accounted for by differences in memory processes. This will be accomplished by incorporating the R/K paradigm (Tulving, 1985, Gardiner, 1988) in face and voice recognition. The aim of this chapter is to explore whether different memory processes are associated with faces and voices — even when overall recognition between faces and voices is matched.

To overcome the disadvantages of using a small stimulus set (Schmidt-Nielsen and Stern, 1985) and to control for any semantic priming effects which may occur with personally familiar items, the test items for this thesis will use famous faces and famous voices belonging to people from the media (Hanley, Smith and Hadfield, 1998). These famous items will be mixed with a selection of non-famous faces and voices. Collectively, these items will be used in chapters 3-6, to assess how effectively participants can discriminate between famous and non-famous items and to establish how successful they are at retrieving person specific information (occupations and names) to the faces and voices that they deem familiar. The selection procedure of the stimuli is outlined in detail in the following chapter (Chapter 2).

## Chapter 2: Selection of Stimuli.

This chapter describes in detail the procedure employed in selecting and creating the stimuli used in the experiments described in Chapters 3, 4, 5 and 6. The section begins with a general description of the celebrity selection process and is subsequently followed up with the guidelines used in selecting and creating the face and voice stimuli. It is worth pointing out that each celebrity and non-famous person used in this thesis acted as their own control across the two modality conditions. For example, a visual clip of Joanna Lumley was presented without any auditory output for the face recognition task and an auditory clip of Joanna Lumley without any visual output was used for the voice recognition task. This in turn allows both face and voice clips to have equivalent durations.

### Part 1 – Selection process.

A total of 45 hours of television programming was recorded from 1999 to 2001 onto video cassettes. The programmes that were recorded included various interview programmes as well as chat show and life style documentaries. This collection generated a total of 79 famous people taken from a wide range of backgrounds; sport, music, television, politics, film and comedy. Approximately 40 non-famous people were also selected from the recordings. In order to verify that the celebrities taken from the television recordings were easily recognizable, a pilot study similar to the one conducted by Schweinberger, Herholz and Stief (1997) was employed.

Each celebrity's name was presented in a list and given to a sample of 12 participants. For each name that was given to them, participants were required to rate their familiarity with the person's name, face and voice. Unlike Schweinberger et al's (1997) procedure, this study used a four point familiarity rating scale rather than a three point scale. The rating scale ranged from 0 = Unfamiliar, 1 = Low Familiarity, 2 = Medium Familiarity and 3 = High Familiarity (Hanley, Smith and Hadfield, 1998).

Each celebrity's mean familiarity rating was calculated across the 12 participants for each of the three name, face and voice ratings. The ratings study produced means (standard deviations) of 2.57 (0.71) for names, 2.32 (0.82) for faces and 1.88 (0.86) for voices.

The familiarity ratings from each of the 79 celebrities were looked at in more detail and those celebrities who had an average familiarity score of less than 1 (Low familiarity) for either the names, faces or voices ratings were discarded. This process eliminated 5 celebrities from the original set. The remaining 74 celebrities were digitally formatted and imported into a multi-media editing program, Final Cut Pro on the Apple Mac to adjust for perceptual and semantic features.

### Part 2a: Faces – controlling for semantic cues.

The traditional approach used to test the face recognition route of Bruce and Young's (1986) model, involved the use of photographic stills of celebrities as test stimuli (Young, McWeeny, Hay and Ellis, 1986). Although such studies have contributed to our understanding of how faces are encoded and subsequently recognized, it is unclear what role if any, movement has in the recognition process. This is a particularly interesting research question to address as in real life we usually encounter moving faces, not static ones.

The past 10 years has seen an increasing trend in researching the role of movement in face recognition. In terms of facial expression processing, it has been observed that movement has a strong contributing role. The most notable case study that looked at movement in facial expression recognition was provided by

Humphreys, Donnelly and Riddoch's (1993) patient H.J.A. H.J.A's ability to correctly judge facial expressions was facilitated by facial movement. When he was presented with static images of facial expressions his performance to correctly make expression judgments was 60% compared to 86% when the facial expressions were presented as moving light-dot displays.

The role of movement has also been investigated within the context of familiar face recognition. Specifically, Lander, Christie and Bruce (1999) and Lander, Bruce and Hill (2001), using the faces of celebrities, have shown that across a range of different degradation conditions, movement leads to better recognition than static images. Lander, Bruce and Hill, (2001) argued that one benefit that movement can offer is that it provides the viewer with information about how the face changes over time and in doing so can provide the viewer with 'characteristic motion signatures' for individual faces (Lander, Bruce and Hill, 2001, p114).

Considering the ecological value of using moving images, the facial stimuli used in the experiments throughout this thesis will depart from the traditional approach of using photographic stills and use moving famous and non-famous faces to test participants' face recognition ability. The type of movement captured for each facial stimulus included both rigid (head nodding and shaking) and non-rigid movement (speech movement and expression changes).

Using famous faces as test stimuli carries the disadvantage that recognition can be made on contextual cues such as clothing and background. This is often controlled for by placing a circular template on the photograph to occlude these non-facial cues (Young, Mcweeny, Hay and Ellis, 1986). This occlusion technique was also administered to the faces used in this thesis. This was accomplished digitally through the use of the cropping tool made available on the Final Cut Pro editing program and is illustrated on the John Major example displayed in figure 2.1a and 2.1b.

Figure 2.1a. Original version of John Major clip with clothing cues.

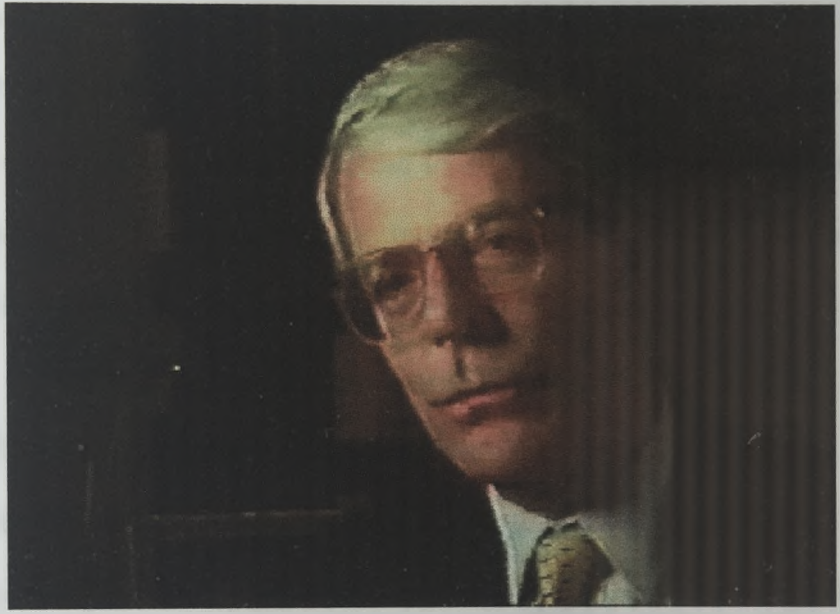
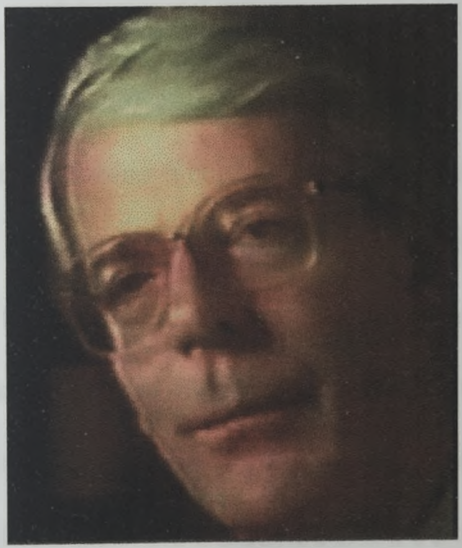


Figure 2.1b. John Major clip with clothing cues removed with cropping tool from Final Cut Pro.



This cropping procedure was applied to all visual clips so that each celebrity and non-famous person was presented to the participant from the neckline upwards. Whilst this cropping tool was very efficient at removing clothing and background cues, it was problematic in cropping those celebrities whose movements were too animated. The cropping tool could only be applied to one fixed position and whilst it occluded the contextual cues such as clothing and background, for some animated celebrities it also occluded their face. This in turn lead to a further reduction in celebrities and non-famous people from the original set. In total, 25 celebrities and 20 non-famous people were discarded.

### Part 2b: Voices – controlling for semantic cues.

To reduce the possibility that participants could adopt selection strategies extraneous to voice recognition, a number of stringent guidelines were followed. The perceptual guidelines outlined by Van Lanker et al. (1985, page 22) in their selection of famous voices were followed in this thesis.

Firstly, each speech sample was carefully selected to be free of any sort of speech trademarks (e.g, Michael Barrymore's 'Awight!') and background noises (studio audience) and identifying sounds (chat show theme tune). Following this procedure a further 6 celebrities were discarded.

Other perceptual guidelines that were adhered to for each voice sample, were taken from Schweinberger, Herholz and Steif's (1997) study. The guidelines emphasized the use of speech samples that were spoken in an emotionally neutral tone and with normal intensity. Furthermore, each sample (famous and non-famous) was edited in Final Cut Pro to synchronize its onset with the onset of a phrase or sentence with the average duration of the voice samples lasting 7 seconds. This is almost 4 times greater than the duration used in tasks 1 and 2 of Van Lancker et al's (1985) study. Additionally, it has been pointed out by Schweinberger, Herholz and Sommer (1997) that voice recognition performance is not likely to improve with longer durations beyond 2 seconds.

The voice modality is perhaps more difficult than the face modality in terms of controlling for contextual cues. It is therefore crucial that the content of each speech sample is neutral in terms of identity. The neutrality of each famous sample was assessed by two means. Firstly, in accordance with the procedure used by Van Lanker et al (1985), each famous sample was presented in a written format and presented to 10 participants. These participants were simply required to guess the famous person from the written samples. From the 43 remaining celebrities, 1 was correctly identified by 30% of the sample and a further 30% of the sample could associate the correct occupational background to two other celebrities. On the basis of these findings, these 3 celebrities were discarded from the selection process.

The second stage of the neutrality assessment followed the procedure adopted by Schweinberger et al (1997). For this assessment, written transcripts of famous and non-famous people were presented in the same fixed random order as those to be used in experiments 1a, 1b and 1c. Participants were given each written format and asked to indicate for each written statement whether they associated it with a famous or non-famous person. They were asked to write the letter F for famous and N for non-famous, respectively.

The results from this pilot study showed that the written texts from famous people were judged to have been spoken by a famous person slightly less frequently than those from non-famous people (0.45% vs. 0.53%). A paired samples t-test showed that the difference was not significant, $t\ (9) = 1.50$, $p > 0.05$. After participants had made their F / N distinction they were asked if they could associate the written text to a particular celebrity. In 20 out of 600 opportunities (10 participants x 60 written formats), participants offered a guess of the target person. However, none of these guesses were correct.

**2.2: Final selection.**

After all of the stimuli had been carefully edited and assessed, the remaining set of items was now 40 famous and 20 non-famous. The familiarity ratings that had been conducted previously on the original 79 famous items were looked at again in more detail. Familiarity ratings were consistently within the high range for faces and were often higher than familiarity ratings made to voices. The selection of the famous stimuli to be used in experiments 1a, 1b and 1c was accomplished by choosing the celebrities with the highest voice ratings. Once this was done, the mean familiarity ratings for names, faces and voices was recalculated after the relevant celebrities had been discarded. The familiarity rating for names was 2.78 (SD = 0.40), for faces 2.56 (SD = 0.55) and for voices 2.13 (SD = 0.61). The total number of experimental trials used in experiments 1a, 1b and 1c was 48 (32 famous and 16 non-famous). Six famous and 4 non-famous items were allocated to the practice trials. Appendix 1 lists the name of the celebrities used in experiments 1a, 1b and 1c.

# Chapter 3: The relationship between familiarity and person identification. Is it the same for faces and voices?

## 3.1: Introduction.

The experiments carried out in this chapter of the thesis will examine the way in which familiarity decisions are made to famous faces and voices and how effectively participants can retrieve person-specific information to faces and voices that they find familiar.

After the initial structural features of the face are processed, the familiarity stage of person recognition is considered to be the precursive stage for identity-specific information retrieval. The initial recognition judgment is usually reinforced by attempting to retrieve more detailed contextual information to the face, such as the person's occupation or where they were encountered previously, for example. The inability to retrieve such information is the hallmark of the familiar-only error observed in person recognition failures (Young et al, 1985a).

Most of the work that has looked at successful retrieval of person-specific information has predominantly used faces as test stimuli (Young, Ellis and Flude, 1988). The use of face stimuli has also been a favourable stimulus in the examination of person recognition errors, both in the laboratory (Hanley and Cowell, 1988) and with neuropsychological cases (Flude, Ellis and Kay, 1986). The wealth of research that has accumulated from the study of successful and unsuccessful retrieval has allowed researchers to propose a three-stage process in the recognition of familiar people.

As reviewed in the chapter 1 (Section 1.6) Bruce and Young's (1986) box and arrow model outlines these sequential stages as familiarity, semantic information and name retrieval. Bruce and Young's (1986) model captures the familiarity stage of person recognition in terms of activation at the level of the Face Recognition Unit (FRU). The 'firing' of the relevant FRU is assessed by the familiarity check role of the cognitive system (Ellis, 1983). This in turn forms the basis for judging the familiarity of a given face. In normal circumstances, the ability to make correct familiarity judgments is a very fast process, with latency studies revealing reaction times of 775 msec in deciding whether a politician's face is familiar (Young, Mcweeny, Hay and Ellis, 1986). However, if the FRU fails to fire or its activation is not deemed strong enough by the cognitive system, then a familiar person will be incorrectly categorized as unfamiliar. Errors at this stage of the facial recognition system has been observed in healthy participants (Young et al, 1985a) aswell as in the neuropsychological literature (Rapcsak, Polster, Comer and Rubens, 1994).

These studies have highlighted the importance of correct familiarity judgments in person recognition memory. However, very little is known about the role of familiarity judgments in voice recognition. This is an important theoretical point that needs to be considered when examining the various person recognition memory models.

## 3.2: Priming in person recognition memory.

As described in the Introduction (see Section 1.8), the Bruce and Young (1986) model was subsequently modified within an interactive activation and competition (IAC) architecture in order to accommodate the neuropsychological behaviour of M.E (de Haan, Young and Newcombe, 1991). This in turn led Burton et al (1990) to propose that familiarity decisions are taken at the level of the PINs – the 'gateway' to person-specific information. All units within the PINs pool are set to a common threshold. Familiarity decisions arise when the signal reaching the PINs exceeds this threshold.

As the PINs also represent the converging point for the structural units (FRUs –for faces, NRUs –for names and VRUs for voices, respectively), then it follows that if a face of a person is found familiar then the person's voice will also be found familiar. Activation from the PINs leads to the store of person-specific information in the Semantic Information Units (SIUs). There is also a common threshold mechanism in place for the SIUs pool, with any unit exceeding this threshold resulting in retrieval of person-specific information.

Non-clinical data in the form of repetition and semantic priming studies has been used to support the architectural integrity of the model. The priming paradigm has been used to investigate the facilitative effects of response to one stimulus following earlier exposure to the same (repetition priming) or related (associated / semantic priming) stimulus. Priming is observed in the form of shorter reaction times, and / or increased accuracy in recognizing previously presented items compared to new stimuli. Whilst the priming paradigm has been largely employed within the verbal recognition domain and has served as an invaluable tool in the development of visual word processing models (e.g. Morton, 1979), parallel experiments have been found in the person recognition literature using faces. For example, Bruce and Valentine (1985) found that the recognition of a familiar face was primed by a previous presentation of the same face, but not by the same person's name. Furthermore, as observed by Ellis, Young and Flude (1990), when participants were required to make expression or gender judgments in both phases of the priming paradigm, no repetition priming occurs on the second presentation. This particular finding seems to indicate that the priming effect is confined to familiarity judgments.

In subsequent studies, Bruce and Valentine (1986) investigated the effectiveness of associative priming using faces. For example, deciding that Stan

Laurel's face is familiar is quicker when it has been immediately preceded by Oliver Hardy's face. These repetition and associative priming effects have been readily accommodated within the box and arrow model (Bruce and Young, 1986). According to Ellis (1992, cited in Ellis, Jones and Mosdell, 1997) the locus of repetition priming is at the FRUs stage, whereas associative priming has been confined to the PIN stage. Considering that familiarity is relocated at the PINs in the IAC model, Burton et al (1990) explain associative priming in terms of excitation among related PINs sharing semantic information. For example, when Oliver Hardy's PIN is activated by the presence of his face this leads to partial activation of Stan Laurel's node. A picture of Stan Laurel's face will be responded to more quickly if it is subsequently presented. Furthermore, since the PIN is viewed as a converging point for all other person recognition routes (NRUs and VRUs), then Stan Laurel will be responded to more quickly if his name or voice is subsequently presented. Repetition priming is generated in the IAC model through the strengthening of the links between FRUs and their corresponding PINs.

A critical factor differentiating the two priming paradigms is the time course each one takes. Associative priming is abolished by any item occurring between the target and prime. Repetition priming on the other hand may remain over long intervals. Additionally, repetition priming is normally a unimodal process whereas associative priming may be multimodal (Ellis et al, 1997).

Ellis et al (1997) investigated repetition priming effects during intra-modal priming (face-face, voice-voice) and inter-modal priming (face-voice, voice-face) across long and short prime-test intervals. For the long interval (10 minutes), repetition priming effects were observed within each modality, so that faces primed previously presented faces and voices primed previously presented voices. There was no evidence of repetition priming across modalities. This latter finding is consistent with both person recognition models. The absence of an inter-modal priming effect fits in with the box and arrow model, as there is neither a direct link between the FRU and VRU stages nor indirect communication via the PINs (Burton et al, 1990). However, the two models differ in that the IAC model *does* allow for inter-modality (face-voice) and (voice-face) priming effects to occur during very short prime-target intervals. This is only possible through the associative priming mechanism of the PINs as illustrated in the Stan Laurel / Oliver Hardy example above. Specifically, when a person's PIN is activated by one modality the activation that is passed within the PIN means that priming can be easily triggered from another modality. If the interval between the person's voice and his/her face is brief and is not interrupted by another item then priming should occur.

When the prime-test trials were brief and uninterrupted (Experiment, 2), inter-modal priming did occur, thus replicating previous cross repetition priming effects observed with names and faces (Calder, 1993 cited by Ellis et al, 1997). However, as pointed out by Ellis et al (1997) this cross-domain priming for faces and voices is of greater biological significance than the name-face effects, thus supporting findings from developmental studies which have shown that the perceptual skills that allow us to identify faces and voices is available early on in life (DeCasper and Fifer, 1980 and Bushnell, Sai and Mullin, 1989). From an evolutionary viewpoint, this cross-domain priming effect supports the notion that faces and voices are inseparably attached to a specific individual, with both types of cues acting as carriers of identity information (Schweinberger, Herholz and Steif, 1997).

### 3.3: The role of the PINs in face and voice recognition.

Since the development of the IAC model in the early nineties, the model has generated a wealth of research in evaluating its cognitive architecture. However, whilst highly insightful, this research has predominantly focused on the 'front' and 'back' end of the model. For example, in its most recent modification, (Burton, Bruce and Hancock, 1999) the 'front' or input end of the model has integrated a Principal Component Analysis (PCA) unit for the FRUs. One function of the PCA unit is to deliver information about the ways in which faces vary.

From the 'back' or output end of the model, the controversial proposition that names should be placed alongside other information in the SIUs has amassed wide research interest (e.g; Cohen, 1990b; Hanley, 1995). As summarized in the Chapter 1 (See Section 1.8), this idea is in stark contrast with the previous proposition made by Bruce and Young (1986). They viewed names as being stored in a separate cognitive stage from other person-specific information.

In light of this research, it is rather surprising to see that familiarity decisions and subsequent retrieval of semantic information, is a topic that has received considerably less attention in the person recognition literature compared to the research that has accumulated for the front and back end aspects of the model. With the exception of Hanley, Smith and Hadfield, (1998), Ellis et al (1997) and Scweinberger et al (1997) most studies that have attempted to look at these issues have primarily used faces and names as test stimuli, with voices receiving little or no attention.

Whilst the IAC model elegantly captures this inter-modality effect in repetition priming and in turn reinforces the notion that the locus of familiarity is at the PINs, Ellis et al (1997) argue that the model is not as robust in accommodating other types of performance measures. For example, Ellis et al (1997) claim that the tip-of-the-tongue state (TOTs) occurred significantly more in response to voices than faces. This person recognition error occurs when the face is associated with some contextual information (e.g., occupation) however the person's name is temporarily inaccessible. Burton, Bruce and Hancock (1999) describe the role of the PINs as 'the locus for familiarity decisions. When any PIN reaches a common activation threshold, familiarity is signaled. This has the implication that the same decision mechanism is used for all person familiarity judgments, regardless of whether they are made to faces, names or other kinds of information' (p.3). With this in mind, it would follow that this type of person recognition error would occur equally often to faces and voices. As argued by Ellis et al (1997), there must be differential strength from the structural links (FRU, VRU) leading up to the PIN stage which in turn would result in differential recognition from faces and voices. If the same decision mechanism is used for familiarity judgments (PIN) for faces and voices then both modalities should lead to full person recognition. The fact that Ellis et al (1997) have shown that TOTs occur more often for voices than faces challenges the assumption of the modality free PINs as originally conceptualized by Burton et al (1999). Hanley, Smith and Hadfield (1998) reported a similar pattern of results when investigating the frequency of the familiar-only error in response to famous faces and voices.

The purpose of the experiments in this chapter is to extend Hanley et al's findings (1998) by using a new set of stimuli and applying a different experimental design. As the experiments in this chapter are based on the procedure employed by Hanley et al (1998) the following sections will describe in detail this procedure and the specific findings related to face and voice recognition performance.

The primary research interest of Hanley et al's (1998) study was to examine how readily Bruce and Young's (1986) model and Burton et al's (1990) model can accommodate the familiar only error. In terms of Bruce and Young's (1986) box and arrow model (see Section 1.6), as familiarity is represented at the structural level (FRUs, VRUs and NRUs) then the familiar only error can arise when the relevant structural component reaches threshold without activating the PIN. This original conceptualization was modified by Burton et al (1990) in their IAC model. Unlike the box and arrow model, the IAC model explicitly defined the role of the PINs by suggesting that they do not contain semantic information about people. Person-specific semantic information is retrieved from the semantic information pool – this stage of the model can only be accessed via the PINs (see section 1.8). By creating this modification, Burton et al (1990) were able to shift the locus of familiarity away from the structural level to the PIN stage of the model. This means that the familiar-only error is explained in the IAC model as resulting from a 'block' between the PINs and the semantic information units (SIUs). As a result of this modification, the model can adequately capture the pattern of performance displayed by de Haan et al's patient (1991), M.E.

This change in locus of familiarity would lead to different predictions of the familiar-only error. If the familiar-only error occurs as a result of a 'block' between the PINs and the SIUs, then this type of error should occur equally often to faces and voices. This is because the connection between the PINs and the SIUs occurs after the two modalities have converged (see Section 1.8). In the case of the box and arrow model (Bruce and Young, 1986), a difference in the number of familiar-only responses across faces and voices could be permitted as familiarity is activated at the level of the structural codes (FRUs, VRUs). This activation may be sufficient to reach threshold, thus triggering a feeling of familiarity, but may not be enough to activate the PINs to retrieve person-specific information.

It is also worth noting that the IAC model can also capture the familiar-only error in terms of a 'block' between the structural codes (FRUs, VRUs) and the PIN. This is because the cascade nature in which the model operates means that if the links leading from the structural codes are relatively weak then there may be enough activation at the PIN to signal familiarity, but not enough activation to pass on through to the semantic information pool.

To see which of these two accounts may be more likely to explain the familiar-only error, Hanley et al (1998) assigned their participants to one of two conditions. For the face condition, participants were presented with a videotape containing 60 clips. There were a total of 40 celebrities used, which were taken from a variety of backgrounds; sport, comedy, music, film , politics and TV personalities. These were randomly interspersed with 20 non-famous people. Participants in the face condition could see each face appear on the television screen. However, the sound was turned off so that the participants could not hear the person's voice.

The voice condition involved the presentation of the same videotape, however participants were seated to the side of the television screen so that they could not see the speaker's face. Each extract was approximately 10 – 15 seconds in duration. Participants in both conditions were required to provide a rating of familiarity which ranged from zero if they found the face unfamiliar and then 1 for low, 2 for medium and 3 for high familiarity, respectively. Furthermore, participants were asked to provide occupational and name details for each face or voice that they deemed familiar.

Hanley et al's (1998) hit and false alarm data showed that participants were much better at discriminating celebrities from foils in response to a face than a voice. Specifically, the face condition produced hit and false alarm rates of 94% and 21%, respectively. For voices, the mean hit rate was 70% and 32% for the false alarm rate.

The critical aspect of their data were revealed in the occupational recall performance. Occupational recall for faces did not prove to be too problematic. The performance measures revealed that if a face was found familiar, then participants could recall appropriate occupational details about the person on 92% of occasions. With voices, the performance measures showed that occupational details were only recalled on 63% of occasions in which the voice was found familiar. These particular findings indicate that not only do participants in the voice condition rate fewer of the celebrities as familiar, but for the ones that they do find familiar they are unable to retrieve occupational details. This in turn demonstrates that there are significantly more familiar-only responses made to voices than faces. Furthermore, there were significantly more familiar-only responses in the voice condition even when participants' subjective familiarity was rated as being very strong. This suggest that participants in the voice condition are poorly calibrated in terms of the relationship between familiarity and person-specific information.

In terms of Bruce and Young's model (1986), it may be the case that there are stronger links between the FRU and the PIN relative to the VRU and the PIN. This in turn would result in fewer occupational details being recalled to familiar voices than familiar faces. This is exactly what was found in Hanley et al's (1998) experiment. Additional support for this interpretation of the data was found in the second phase of the testing procedure. For those participants who were unable to recall occupational details to a face or voice that they found familiar, they were subsequently cued with either the same modality or a different modality of the person. For example, if the participant was initially presented with the person's face they could either be cued again with the same clip of the person's face or with the person's voice. During this cueing stage of the experiment, it was found that participants who were initially in a familiar-only state for voices found faces particularly beneficial in resolving the familiar-only state. For participants that were in a familiar-only state for faces, presenting a voice cue did not prove to be as successful in resolving the familiar-only state as when a face followed a voice cue. This asymmetric cueing effect fits in neatly with the prediction that there are stronger links between the FRU and PIN than between the VRU and PIN. If the familiar-only error occurred at the pathway connecting the PINs and the SIUs, then a new modality cue would not be any more beneficial than a same modality cue. This is because the new modality cue does not provide an alternative means to access semantic information about a person.

Hanley et al (1998) also assert that the IAC model can readily capture their data. As described previously, the cascade architecture of the model means that weak activation at the level of VRU will lead to reduced activation at the PIN – irrespective of the strength between the VRU-PIN link. A consequence of this is that the PIN may receive sufficient activation from the VRU to signal familiarity, however the strength of this activation is too weak to allow access to semantic information.

The only aspect of Hanley et al's (1998) data that is difficult to reconcile with the IAC model is the finding that there were significantly more familiar-only responses made to voices even when the familiarity of the voice was deemed to be very strong. If familiarity decisions are produced by activating the PINs then it would be reasonable to assume that the stronger the feeling of familiarity the stronger the activation at the PINs which in turn would facilitate retrieval of person specific information. However, as it has already been pointed out, familiar-only responses still occurred even when familiarity was reported to be very strong. As the current IAC model stands there is no explicit implementation of subjective strength of familiarity. Therefore, participants find a face or voice familiar if activation levels at the PIN exceeds a threshold. There is no mechanism that signals degree of familiarity to the perceiver. Thus, the model is immune to data of this kind. Overall the findings from Hanley et al's (1998) study not only demonstrates superior recognition ability to faces than voices, but also that there is a stronger relationship between familiarity and identification (calibration) in faces than for voices.

At this point, it is worth noting that Hanley et al's (1998) study employed a between subjects design. This is an important point to highlight as it might provide some clues as to why participants in the voice condition show poor calibration between varying levels of familiarity and occupational recall. As Hanley et al (1998) used a between subjects design, there is a possibility that participants used a different criteria for familiarity decisions in the voice and face conditions. Making familiarity decisions to voices alone may not have provided a sufficient enough context for participants in the voice condition. The role of context on memory performance in terms of calibration has been looked at in detail in the confidence-accuracy (C-A) literature in the form of the mixed question effect (Hollins and Perfect, 1997).

### 3.4: The Mixed question effect.

The confidence-accuracy (C-A) literature has demonstrated moderate to strong calibration between confidence and accuracy for general knowledge (see Perfect, Watson and Wagstaff, 1993), but weak and non-significant correlations for eyewitness memories (see Bothwell, Brigham and Deffenbacher, 1987 and Smith, Kassin and Ellsworth, 1989). One account that has been put forward to explain the lack of C-A correlation in eyewitness memory that is otherwise present in general knowledge memory, is that the latter provides the individual with an insight into their own strengths and weakness whilst the former does not. General knowledge memory enables participants to cross-reference their potential answer through assessing their own areas of expertise with the subject and in turn assess their relative standing compared to others. This process has been referred to by Hollins and Perfect (1997) as 'confidence-calibration-through experience' (p. 215). This allows participants in the general knowledge memory tests to correctly assign their memories to the relevant category on the confidence response scale. In contrast, eyewitness memory does not provide the participant with an insight into their areas of expertise (how good am I at recognizing someone's clothes ?) nor does it provide the participant with an insight into their relative standing compared to others. One consequence of this is that the response scale will be used more subjectively than in the general knowledge test, which in turn will lead to low and non-significant C-A correlations for eyewitness memory.

Hollins and Perfect (1997) sought to improve the C-A correlation in eyewitness memory by intermixing general knowledge questions with eyewitness questions to form a single memory test. The rationale behind this was to see whether providing participants with an adequate context in eyewitness memory could improve the overall C-A correlation for eyewitness memory. The context that was supplied to the participants was in actual fact their own confidence during general knowledge memory test items. This means that the initial judgement that would have been employed by the participant in the eyewitness memory task; 'how confident am I about this fact?' now becomes 'how confident am I about this fact compared to how confident I said I was about general knowledge facts?' (Hollins and Perfect, 1997, p.208). By providing participants with such a context it was anticipated that participants would become better calibrated in terms of their confidence for eyewitness memory items than when they were not provided with this context.

Hollins and Perfect (1997) arranged their general knowledge and eyewitness test materials in one of two ways. Participants either had to complete the two tests in blocks; eyewitness test items followed by general knowledge test items or they had to complete a 'mixed' test whereby eyewitness and general knowledge items alternated for each question. Additionally, both versions were administered in the form of a recognition and recall test.

Hollins and Perfect (1997) observed consistent effects of questioning mixing, in which significant C-A correlations emerged for eyewitness memory when eyewitness memory questions were intermixed with general knowledge questions than when they presented in separate blocks. It therefore appears that one way in which participants can become successfully calibrated is by comparing their memories with a useful context. In the case of Hollins and Perfect (1997) the participants' own confidence on a different memory task (general knowledge) was sufficient in helping them achieve better calibration in the eyewitness task.

This subtle, yet effective change in experimental procedure has also been observed in the perceptual domain, such as categorical perception. Categorical perception is associated with better discrimination across category boundaries than for equivalently spaced stimuli within the same boundary (Pilling, Wiggett, Özgen and Davies, 2003). In a series of experiments using colour stimuli and facial expressions, Roberson and Davidoff (2000) found that the advantage of categorical perception was removed when participants were given a verbal interference task. The fact that categorical perception was maintained on trials that used visual interference lead these authors to conclude that the advantage of cross-category judgements over within-category judgements arises because participants actively attempt to name stimuli. When this is prevented through a verbal interference task, categorical perception is abolished.

However, as noted by Pilling et al (2003), Roberson and Davidoff's (2000) interference tasks were administered in a blocked order. The implication of this, Pilling et al (2003) argue, is that the interference tasks were predictable and therefore participants may have used different strategies in the different interference conditions in order to avoid the effects of interference. Pilling et al (2003) tested this hypothesis by administering the interference trials randomly rather than through blocked presentation. When this aspect of the procedure was changed, Pilling et al (2003) found that categorical perception was no longer abolished after verbal interference. If the interference trials are predictable (blocked), then this directly affects how the first task is performed. For example, participants may switch between two different strategies, so that when they know that verbal interference will follow the encoding task, they may purposely use visual codes to encode the stimuli, whereas visual codes may be used if a verbal interference task is expected.

Given the findings obtained by Hollins and Perfect (1997) and Pilling et al (2003), the experiments in this chapter will aim to address whether changes in experimental design can affect the level of calibration in familiarity decisions and occupational recall for faces and voices.

### 3.5: Rationale.

There are two alternative ways in which this familiar-only bias in voices can be studied. One approach adopted by Hanley and Turner (2000) involved comparing occupational performance between faces and voices when overall recognition between the two modalities is equated. Hanley and Turner (2000) argued that the familiar-only bias in the voice condition in their earlier study may have come about through voices being associated with lower levels of familiarity than faces. To test this particular hypothesis, Hanley and Turner (2000) carried out their original study with the inclusion of a third modality condition. This new condition comprised of the original faces being presented in a blurred format. This particular degradation procedure produced hit and false alarm rates that were similar to the voice condition (see table 1, Hanley and Turner, 2000).

With this modification, there were just as many familiar only responses in the (blurred) face condition as in the voice condition. Furthermore, comparisons between the two face conditions showed that there were fewer occupations retrieved to faces found familiar in the blurred face condition than in the standard face condition. Since the same set of faces were used in both the standard face and blurred face conditions, this particular finding would suggest the criteria participants use to assess their levels of familiarity varies according to overall level of performance. In other words, a 'highly' familiar face in the standard face condition does not appear to mean the same thing as 'highly' familiar rating in the blurred face and voice conditions. Hanley and Turner (2000) added that the criteria participants use appears to differ when overall levels of familiarity are much lower and the ability to distinguish target items from distractors is more difficult. This study by Hanley and Turner (2000), which also used a between-subject design, demonstrated that matching overall performance can provide an insight into the familiar-only bias in voice recognition.

The series of experiments reported here, attempts to approach this issue of criteria differences between faces and voices through varying the experimental design. As Hanley et al (1998) observed more familiar-only responses to familiar voices relative to familiar faces by using a between-subjects design, the first Experiment (1a) aimed to replicate this observation with a new selection of faces and voices. If participants in the two modalities are using a different criteria to make their familiarity judgements, then by administering the experiment in within-subjects design, Experiments (1b and 1c) should provide participants with an adequate context to base their familiarity judgements. What is of specific interest in this study is whether participants can use their familiarity judgements for faces to provide a sufficient context for their familiarity judgements for voices in a within-subjects experiment. If this context does prove to be beneficial in helping participants assign the correct familiarity rating for voices, then differences between the two modalities for occupational recall should not be found.

## Method.

### *Experiment 1a.*

### Participants.

Fifty-four participants who were student and staff members from the University of Essex took part in the experiment. The age range of participants was from 18 to 57 years, with a mean age of 35.4 years (31 female and 23 male). All participants were British and had normal to corrected-to-normal vision and normal hearing. All participants were paid £2 for their participation. Four participants were discarded because they failed to follow the instructions.

### Apparatus.

An Apple Macintosh computer was used to present the stimuli to the participants. Presentation was via the QuickTime Player software programme.

### Stimuli.

The outline of the stimuli selection procedure is presented in detail in Part 2.1.a and 2.1.b of Chapter 2. Overall there were 38 famous clips and 20 non-famous clips used in the experiment. For the experimental trials, 32 famous clips and 16 non-famous clips were used. The remainder were allocated to the practice trial (6 famous and 4 non-famous clips). The clips containing the famous people comprised of approximately equal numbers from the following media backgrounds; politics, film, TV, music, comedy and sport. The full listing can be found in the Appendix 1. The duration of each clip was approximately 7 seconds. The order of the 48 clips on the QuickTime program was randomised.

**Design.**

Following on from Hanley et al's (1998) study, this experiment used a between subjects design. Twenty-five participants were randomly allocated to the face condition and twenty-five participants were randomly allocated to the voice condition. A number of different dependent measures were taken in the assessment of voice and face recognition performance. The dependent measures in this experiment included the number of hits and false alarms made in both modality conditions. A hit is achieved when the participant assigns a familiarity rating between 1-3 to a famous clip (face or voice). A false alarm is achieved when the participant assigns a familiarity rating between 1-3 to a non-famous clip (face or voice). In this case, the correct response would be a zero rating to denote that the clip is unfamiliar. The hit rates to faces and voices shall also be looked at for each of the three levels of familiarity. Participants' ability to retrieve the occupational details and the name to the face or voice that they deemed familiar at each of the three levels of familiarity will also be examined.

**Procedure.**

The participants were tested individually in a small testing booth in the Department of Psychology. Presentation for visual stimuli (face condition) and auditory stimuli (voice condition) was via the QuickTime Player on the Apple Mac. The 48 experimental trials (32 famous and 16 non-famous) were presented in a fixed random order.

In the face condition, the participants were seated approximately 30cm in front of the computer monitor and each face appeared on the computer monitor one at a time. The size of the QuickTime screen measured approximately 14 inches along the diagonal. In the voice condition, participants were also seated 30cm in front of the monitor, except there was no visual output. Here participants listened to the voices via a set of Sennheiser headphones. The presentation rate of each extract was regulated by a key press on the keyboard from the experimenter.

Participants were instructed to decide whether they found the face or voice extract familiar or unfamiliar. The degree of familiarity was classified by using the same rating scale as used by Hanley, Smith and Hadfield (1998), in which '0' was used to denote unfamiliarity and '1' was used to signify low familiarity, '2' for medium and '3' for high familiarity. If the participant found the face or voice familiar, they were also asked if they could recall the person's occupational details and name. If a participant gave a vague description, they were prompted to provide additional information. For example, in response to seeing or hearing Ainsley Harriot, they were prompted to give more detailed information beyond the general description such as 'TV presenter'. For the Ainsley Harriot example, the following information was probed for: 'TV presenter, cookery programmes, Ready, Steady, Cook'. If they could not retrieve any occupational / biographical information to a face or voice that they found familiar, they were specifically requested not to guess the occupational details and to respond by saying 'familiar only'.

The participants were required to give a spoken response, which was recorded on a response sheet by the experimenter. Each clip was presented only once and after each extract, the program was paused whilst the participant gave their response. The participants were requested to respond as quickly as they could. The participants were only given feedback after the practice trial and at the end of the experiment, at which point they were thanked for their participation and de-briefed.

## Results.

Table 3.1a provides the mean scores for overall recognition performance across the two modalities. The mean number of hits is out of a total of 32 and the false alarms are out of a total of 16 for each condition.

### *Overall recognition performance.*

Table 3.1.a. Means and standard deviations (in parentheses) for the face and voice conditions.

| | Faces | Voices |
|---|---|---|
| Hits | 25.2<br>(4.9) | 17.8<br>(5.5) |
| False alarms | 2.6<br>(2.2) | 3.6<br>(2.6) |
| Number of occupations correct | 21.7<br>(5.6) | 8.9<br>(4.3) |
| Number of names correct | 17.3<br>(7.8) | 7.4<br>(4.1) |
| Percentage of occupations recalled conditionalized on familiarity | 82.4<br>(20.5) | 50.2<br>(17.3) |
| Percentage of names recalled conditionalized on familiarity | 62.3<br>(28.3) | 40.6<br>(17.2) |

An unrelated samples t-test was applied on each of the performance measures. There were significantly more hits made in response to faces than voices, t (48) = 4.94,

$p<0.05$. However the false alarm measures were non significant between the two modalities, t (48) = 1.534, $p> 0.05$. This indicates that even though participants were better at recognizing celebrities from their faces than their voices, the number of times an unfamiliar clip was misidentified as a familiar item was the same for both of the modalities. There were significantly more occupations recalled to faces than voices, t (48) = 8.69, $p <0.05$ and there were significantly more names recalled to faces than voices, t (48) = 5.60, $p <0.05$. The occupation and name data were analysed in terms of the proportion of occupations and names recalled as a function of overall familiarity. This showed that significantly more occupations were recalled to faces than voices, t (48) = 5.27, $p < 0.05$. Similarly, the proportion of names recalled as a function of overall familiarity was significantly greater for faces than voices, t (48) = 3.25, $p< 0.05$.

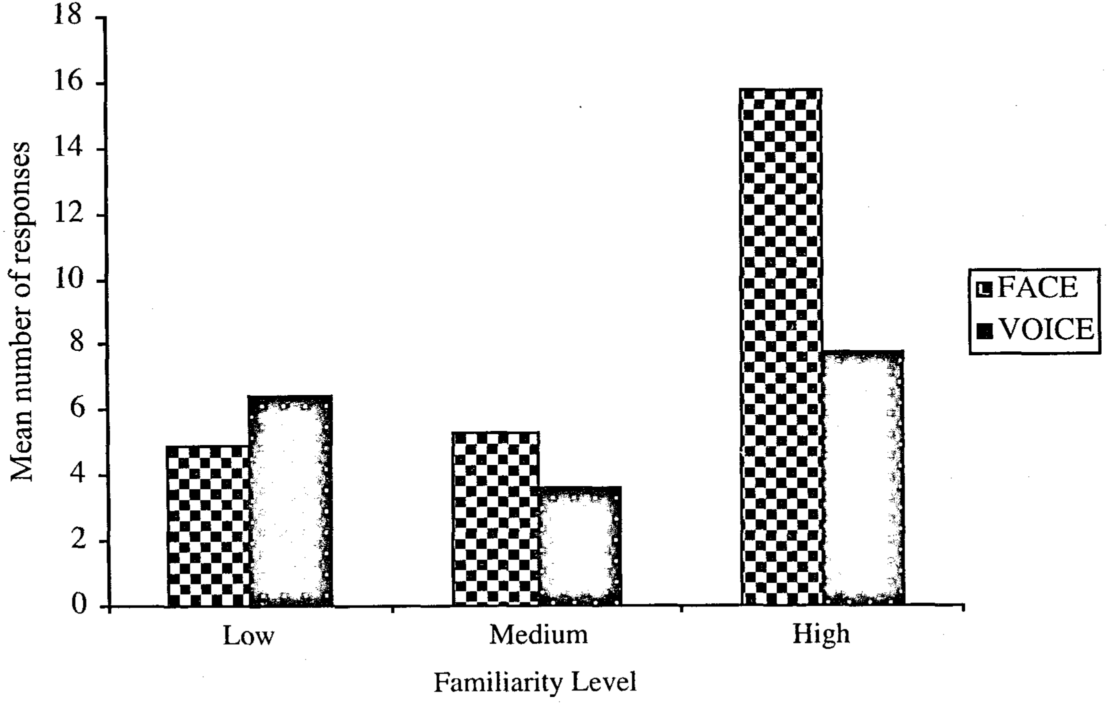
### *Signal Detection Measures.*

Signal detection measures were calculated for the hit and false alarm data of both conditions. Measures of sensitivity and bias were calculated using A' and B"D, respectively. This index of sensitivity (A') is analogous to d' measures (Valentine, Chiroro and Dixon, 1995) Furthermore, A' can be calculated when hit or false alarm rates are 1 or 0. Values of A' can vary between 0 and 1.00, with values approaching 1.00 indicating perfect discrimination of targets (famous clips) from foils (non-famous clips). Values around the 0.50 region, indicate chance performance. The values for the bias measure B'D can vary between –1.00 (extreme liberal response) and +1.00 (extreme conservative response) a value of zero represents neutral bias. In cases where the false alarm rate exceeded the hit rate, Aaronson and Watts' (1987) modified formula for calculating A' was used. For the calculations of B" D, the formula recommended by Donaldson (1996) was used (see Appendix 2).

The mean A' value for the face condition was 0.88 and 0.74 for the voice condition. Both values were statistically above chance ($p < 0.05$). However, the level of discrimination was greater in the face condition than in the voice condition ($p < 0.05$). The degree of bias, as measured by B" D was 0.06 for faces and 0.38 for voices. The degree of bias for both conditions was in the conservative direction. However, the degree of bias did not significantly differ from neutral ($p=.5523$) in the face condition. The performance in the voice condition was significantly more conservative than in the face condition ($p < 0.05$). The fact that voices tended to elicit more conservative responding may explain why the false alarm rate for voices did not differ from faces.

The hit rate data were analysed in more detail by looking at the number of correct familiarity decisions at each level of familiarity for both modalities. This was examined by a 3x2 mixed analysis of variance. Such an analysis was considered appropriate, even though familiarity is not in the traditional sense a true independent variable. There was a main effect of modality, $F(1, 48) = 40.02$, $p<0.05$ and familiarity $F(2, 96) = 35.37$, $p <0.05$ and there was also a familiarity by modality interaction $F(2, 96) = 13.51$, $p < 0.05$. Tukey tests showed that the number of responses at the low familiarity level did not differ between faces and voices. However, there were significantly more medium and high familiarity ratings given to faces than voices ($p < 0.05$). Tukey tests also showed that for faces there were more responses made at high familiarity ratings than low familiarity ratings ($p < 0.05$). Additionally, there were more responses to high familiarity ratings than medium familiarity ratings ($p < 0.05$). For voices, the pattern of responding was slightly different, with only significant differences found between medium and high levels of familiarity ($p < 0.05$). This pattern of results are displayed in figure 3.1a.

Figure 3.1.a. Mean number of responses at each level of familiarity for faces and voices.



This pattern of results fits in nicely with the performance measures presented earlier. Table 3.1a showed that there were significantly more occupations and names recalled to familiar faces than voices. This can be explained in the following way. Figure 3.1a shows that there are significantly more medium and high familiarity responses to faces than voices. The obvious implication is that items that are given higher ratings will facilitate recall than items that are given lower ratings. Because faces are associated with higher levels of familiarity this will lead to better recall of occupations and names than voices which are not associated with high levels of familiarity. However this is not the whole story (see next section).

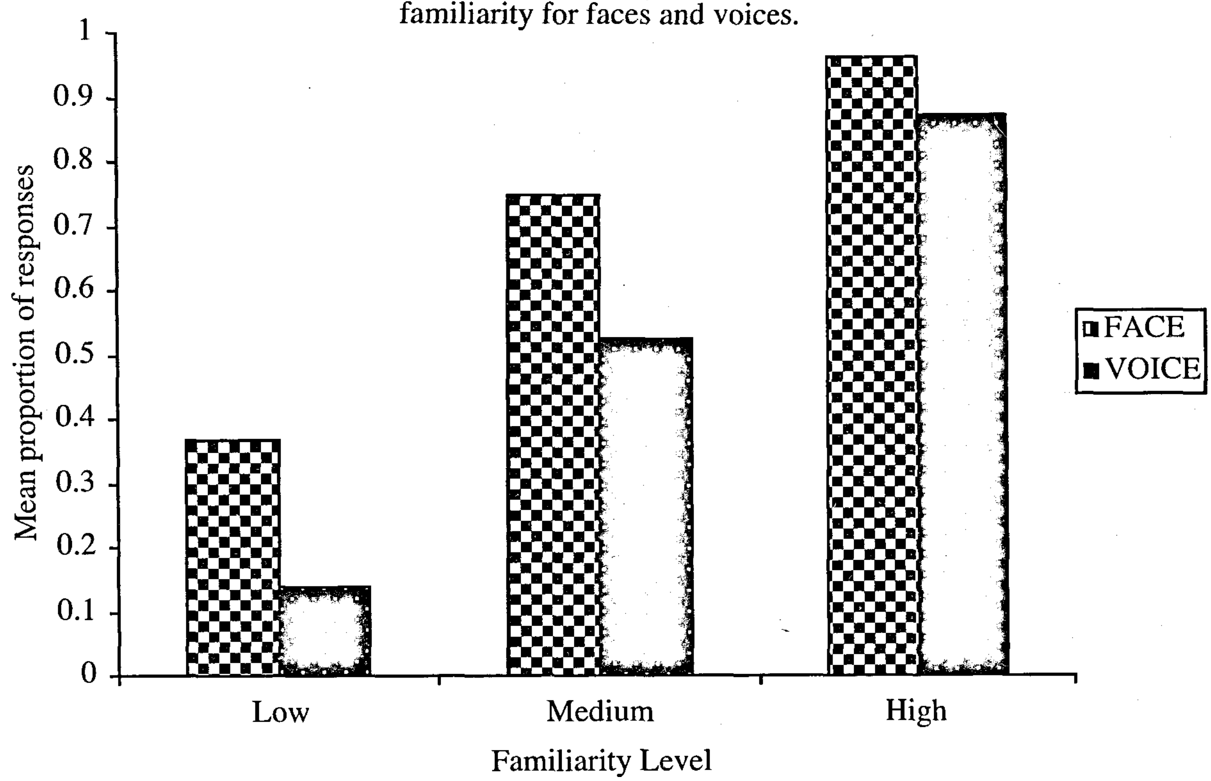
## *Retrieval of person-specific identity information.*

### *Occupational recall.*

A strict marking criterion was applied in analysing participants' occupational responses to faces and voices. Participants were always prompted to provide as much information as possible, as in the Ainsley Harriott example given previously. When participants did give a vague description, even after prompting, their response was classified as a familiar only. Following on from Hanley and Cowell (1988), participants' ability to retrieve occupational details to faces and voices was looked at as a function of familiarity. As pointed out by Hanley and Cowell (1988) a participant may fail to recall a particular piece of person-specific information not because of any retrieval failure per se, but because the biographical details of the celebrity are not stored in the participant's memory. When biographical retrieval is conditionalized on each level of familiarity, then as Hanley and Cowell (1988) point out, all failures to retrieve biographical information should be genuine retrieval failures. This measure will be used as an index of the familiar-only error.

A 3x2 mixed analysis of variance examined the participants' ability to recall occupational details in response to a face and a voice at each of the three levels of familiarity. The unit of measurement was the proportion of occupations correctly recalled conditionalized on each level of familiarity. There was a main effect of modality by participants, $F(1, 48) = 12.75$, $p < 0.05$ and a main effect of familiarity, $F(2, 96) = 109.04$, $p < 0.05$, but the interaction was non-significant $F(2, 96) = 1.46$, $p > 0.05$. A main effect of modality was also observed by items, $F(1, 62) = 24.64$, $p < 0.05$ . The occupational recall data for the two modalities is displayed in figure 3.1.b.

Figure 3.1.b. Mean proportion of occupations recalled at each level of familiarity for faces and voices.

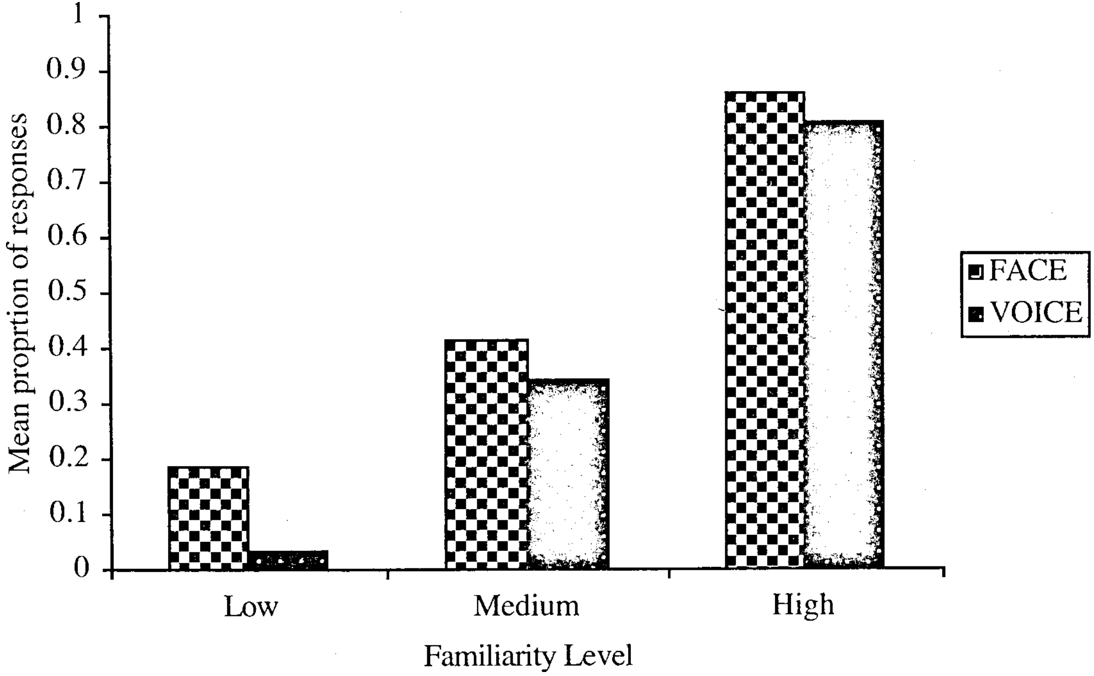


As can be seen there are more occupations recalled to familiar faces than familiar voices. Hence it is not possible to explain the superior recall of occupations to familiar faces than voices simply in terms of higher familiarity of faces than voices. Even when familiarity is equated, recall of occupations is easier for faces than voices.

*Name recall.*

A 3x2 mixed analysis of variance looked at participants' ability to recall the name to a familiar face and familiar voice. The unit of measurement was the proportion of names correctly recalled conditionalized on each level of familiarity. There was only a main effect of familiarity, F (2, 96) = 139.61, $p < 0.05$. Surprisingly, there was no main effect of modality, F (1, 48) = 2.84, p =0.09. The interaction was not significant, F (2, 96) = 0.72, p = 0.484. This pattern of performance can be viewed in figure 3.1c. The data from this performance measures show that the ability to retrieve a name is the same for both familiar faces and voices.

Figure 3.1.c. Mean proportion of names recalled at each level of familiarity for faces and voices.



**Discussion.**

Even when new stimuli are used, the finding established earlier by Hanley, Smith and Hadfield (1998), that there are more occupations recalled to familiar faces than familiar voices is replicated in this experiment. The next experiment will test the hypothesis whether these findings are due to participants using different criteria in assigning their familiarity judgements to faces and voices. By altering the experimental design to a within-subjects design, it was anticipated that familiarity judgements made to faces would provide a context for participants when they are required to make familiarity judgements to voices. By alternating between making familiarity decisions to faces and voices, familiarity decisions should be on the same scale. In a between-subjects design, this would have taken the form of 'how familiar do I find this voice?'. In a within-subjects design, this now becomes 'how familiar do I find this voice compared to how familiar my decisions have been about faces?' This in turn would enforce participants to use the familiarity scale more consistently, which in turn may reduce the overall modality effect for occupational recall that is seen in figure 3.1.b.

## Method.

### *Experiment 1b.*

### Participants.

Thirty-one participants, the majority of whom were undergraduate students at University of Essex, took part in the experiment. The age range of the participants was 18 to 46 years, with a mean age of 31.08 years (22 female and 8 male). All participants were British and had normal to corrected-to-normal vision and normal hearing. All participants were paid £2 for their participation. One participant was discarded because they failed to follow the instructions.

### Apparatus.

Same as in Experiment 1a.

### Stimuli.

Same as in Experiment 1a.

### Design.

This experiment used a within subjects design with all participants exposed to faces and voices. The dependent measures in this experiment were the same as the previous one. They included the number of hits and false alarms made in both conditions. Also the number of responses made at each level of familiarity for both conditions. The correct recall of the celebrity's occupation and name in response to a face or a voice when conditionalized on familiarity was also looked at.

**Procedure.**

The procedure for this experiment was essentially the same as in the previous experiment. The critical difference however was in the administration of the stimuli. Each famous and each non-famous item was presented as a face to half of the participants and as a voice to the other half. The 48 experimental trials were given in a series of 12 blocks with 4 faces/voices per block, alternating between faces and voices in each successive block. The order of trials within the blocks were arranged in a fixed random order. The 48 experimental trials consisted of 32 famous clips and 16 non-famous clips. As participants were exposed to both face and voice modality conditions, the number of famous faces and voices used in this experiment was 16 for each modality. The number of non-famous faces and voices was 8 for each modality.

## Results.

Table 3.2a provides the overall recognition performance across the two modalities. The mean number of hits is out of a total of 16 and the false alarms are out of a total of 8 for each condition.

*Overall recognition performance.*

Table 3.2.a. Means and standard deviations (in parentheses) for the face and voice conditions.

| | Faces | Voices |
|---|---|---|
| Hits | 13.5<br>(2.5) | 9.2<br>(3.3) |
| False alarms | 1.77<br>(1.5) | 2.87<br>(2.1) |
| Number of occupations correct | 11.4<br>(3.8) | 4.3<br>(2.2) |
| Number of names correct | 10.3<br>(4.1) | 3.6<br>(2.2) |
| Percentage of occupations recalled conditionalized on familiarity | 82.2<br>(19.3) | 46.4<br>(20.8) |
| Percentage of names recalled conditionalized on familiarity | 73.8<br>(23.3) | 37.1<br>(22.5) |

A series of related samples t-test were carried out on each performance measure. Performance was better overall in response to faces than voices. There were significantly more celebrities recognized in response to faces than voices, t (29) = 4.27, $p < 0.05$. There were more false alarms given to voices than faces, t (29) = 2.82, $p < 0.05$, thus showing that there were significantly more instances when an unfamiliar foil was misidentified as a familiar item to voices than faces. There were significantly more occupations recalled to faces than voices, t (29) = 10.16, $p < 0.05$ and there were significantly more names recalled to faces than voices, t (29) = 9.1, $p < 0.05$. Finally, the proportion of occupations conditionalized on overall familiarity was significantly greater for faces than voices, t (29) = 8.40, $p < 0.05$. There were also significantly more names conditionalized on familiarity for faces than voices, t (29) = 7.36, $p < 0.05$.

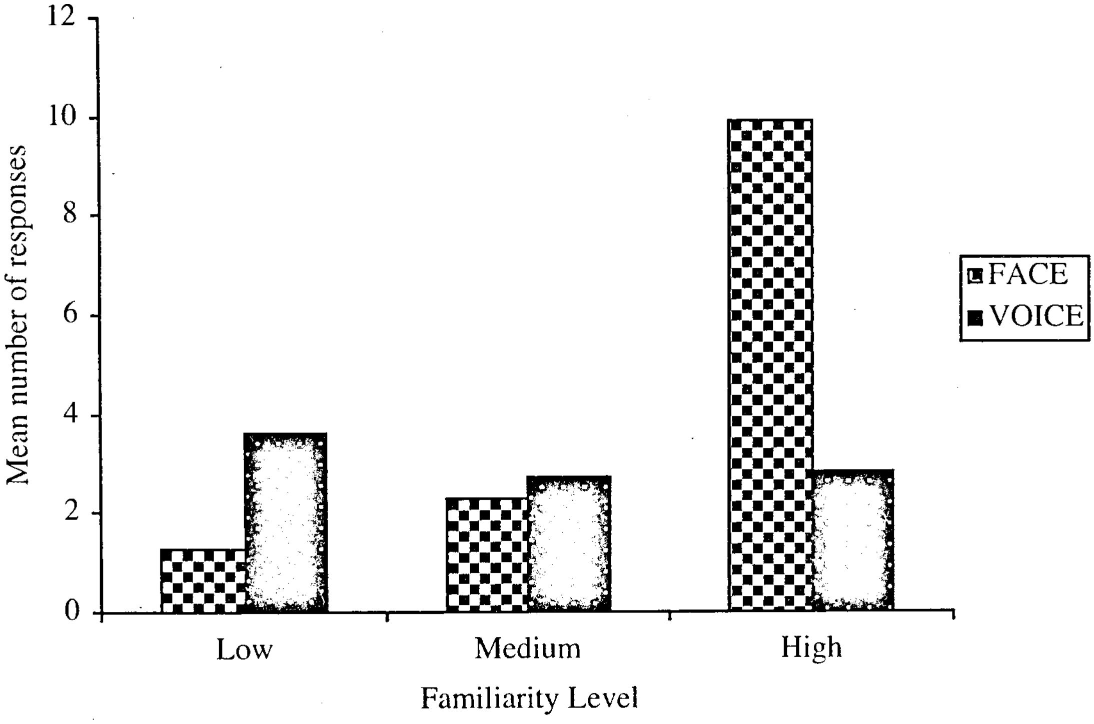
### *Signal Detection Measures.*

The method of calculating the signal detection measures was the same as that applied in experiment 1a. The mean A' value for the face condition was 0.84 and 0.68 for the voice condition. Both values were statistically above chance ($p < 0.05$). However, the level of discrimination was statistically greater in the face condition than in the voice condition ($p < 0.05$). The degree of bias as measured by B" D was –0.20 for faces and -0.007 for voices. Whilst participants adopted a liberal response bias in the face condition ($p<0.05$), participants in the voice condition showed a response bias that did not statistically differ from neutral ($p=0.65$). The overall difference in bias between the two modalities just failed to reach significance ($p=0.0547$).

The number of correct familiarity decisions at each level of familiarity to faces and voices was examined by a 3x2 within subjects analysis of variance. There was a main effect of familiarity, F (2, 58) = 28.47, $p < 0.05$ and a main effect of modality, F (1, 29) = 44.90, $p < 0.05$. There was also a significant familiarity by modality interaction, F (2, 58) = 58.51, $p < 0.05$. Tukey tests showed that there were significantly more high familiarity ratings given to faces than voices. Moreover, voices tended to be associated significantly more with low levels of familiarity than faces ($p < 0.05$). These patterns of results are displayed in figure 3.2.a. This pattern of results may reflect floor effects in the voice condition and ceiling effects in the face condition.

Once again this pattern of results can explain the performance measures in table 3.2.a. Here we can see again that faces were associated significantly more with high levels of familiarity than voices, which were associated more often with low levels of familiarity. Given the fact that high levels of familiarity lead to better recall, then it is no surprise that there are significantly more occupations and names retrieved to faces than voices.

Figure 3.2.a. Mean number of responses at each level of familiarity for faces and voices.

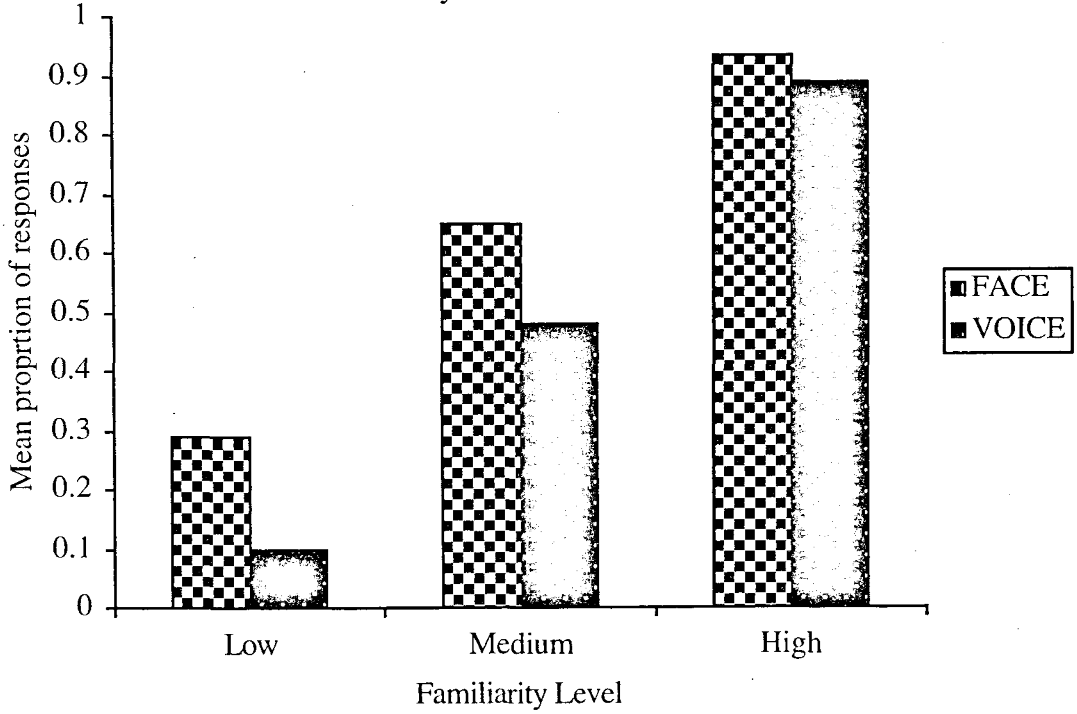


### *Retrieval of person-specific identity information.*

*Occupational Recall.*

Once again, the same strict marking criteria were applied when analysing participants' occupational responses. A 3x2 within subjects analysis of variance examined the proportion of occupations recalled to faces and voices conditionalized at each level of familiarity. There was a main effect of familiarity, $F\ (2, 58) = 89.40$, $p < 0.05$ and a main effect of modality, $F\ (1, 29) = 12.71$, $p < 0.05$. The familiarity by modality interaction just failed to reach significance, $F\ (2, 58) = 3.09$, $p = 0.05$. These results are displayed in figure 3.2b. A further analysis by items also showed a main effect of modality, $F\ (1, 31) = 27.69$, $p < 0.05$ for occupational recall.

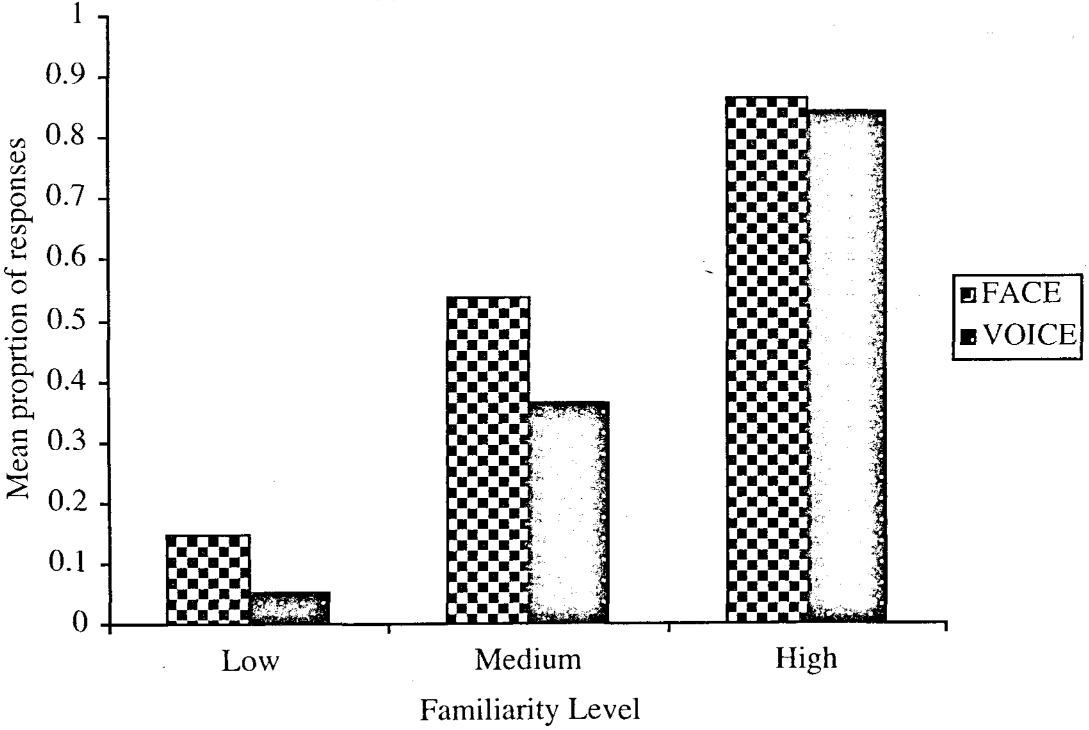
Figure 3.2.b. Mean proportion of occupations recalled at each level of familiarity for faces and voices.



*Name recall.*

A 3x2 within subjects analysis of variance looked at the proportion of names recalled conditionalized at each level of familiarity for faces and voices. There was a main effect of familiarity F (2, 58) = 104.20, p< 0.05 and also a main effect of modality F 1, 29) = 6.14, p < 0.05, but no interaction, F (2, 58) = 2.73, p = 0.07. The effect of modality shows that names are more likely to be recalled to faces than voices even when level of familiarity is equated (Figure 3.2c).

Figure 3.2.c. Mean proportion of names recalled at each level of familiarity for faces and voices.



## Discussion

The aim of this second experiment was to examine whether providing participants with their own familiarity judgements from one modality (faces) will help them assess their familiarity judgements to a different modality (voices). Under this manipulation it was predicted that the familiarity decisions used for faces and voices should be on the same scale. This in turn would encourage participants to use their familiarity judgments more consistently to faces and voices and remove the modality difference observed in the retrieval of biographical information to familiar faces and voice in

Experiment 1a. However, this particular manipulation made no difference. Familiar faces deemed familiar continued to be associated with greater person-specific information than voices (figure 3.2b and 3.2c).

Participants in this experiment were required to alternate between faces and voices after every fourth trial. Under these circumstances the opportunity for calibration may have been reduced compared to an experiment where alternation occurred on every trial. Since Hollins and Perfect's (1997) test materials alternated between general knowledge / eyewitness items on every trial, experiment 1b was carried out again with the alternation of trials modified. Following on from Hollins and Perfect's (1997) procedure, participants in experiment 1c alternated between faces and voices on every trial.

## Method.

*Experiment 1c.*

### Participants.

Thirty participants, the majority of whom were undergraduate students at University of Essex, took part in the experiment. The age range of the participants was 18 to 48 years, with a mean age of 27.2 years (24 female and 6 male). All participants were British and had normal to corrected-to-normal vision and normal hearing. All participants were paid £2.50 for their participation.

### Apparatus.

Same as previous experiments.

### Stimuli.

Same as previous experiments.

### Design.

This experiment used a within subjects design with all participants exposed to faces and voices. The dependent measures in this experiment were the same as in the previous experiments.

### Procedure.

The procedure for this experiment was essentially the same as in experiment 1b. Each famous and each non-famous item was presented as a face to half of the participants and as a voice to the other half. The 48 trials alternated between faces and voices. The experimental trials were arranged in a fixed random order. The 48 experimental trials consisted of 32 famous clips and 16 non-famous clips. As participants were exposed to both face and voice modality conditions, the number of famous faces and voices used in this experiment was 16 for each modality. The number of non-famous faces and voices was 8 for each modality.

## Results.

Table 3.3.a provides the overall recognition performance across the two modalities. The mean number of hits is out of a total of 16 and the false alarms are out of a total of 8 for each condition.

### *Overall recognition performance.*

Table 3.3.a. Means and standard deviations (in parentheses) for the face and voice conditions.

| | Faces | Voices |
|---|---|---|
| Hits | 13.5<br>(1.4) | 9.5<br>(2.7) |
| False alarms | 1.2<br>(1.3) | 2.8<br>(1.9) |
| Number of occupations correct | 10.1<br>(2.9) | 3.5<br>(2.9) |
| Number of names correct | 9.1<br>(2.8) | 3.1<br>(2.9) |
| Percentage of occupations recalled conditionalized on familiarity | 74.0<br>(19.4) | 35.1<br>(26.1) |
| Percentage of names recalled conditionalized on familiarity | 66.6<br>(19.4) | 30.7<br>(25.3) |

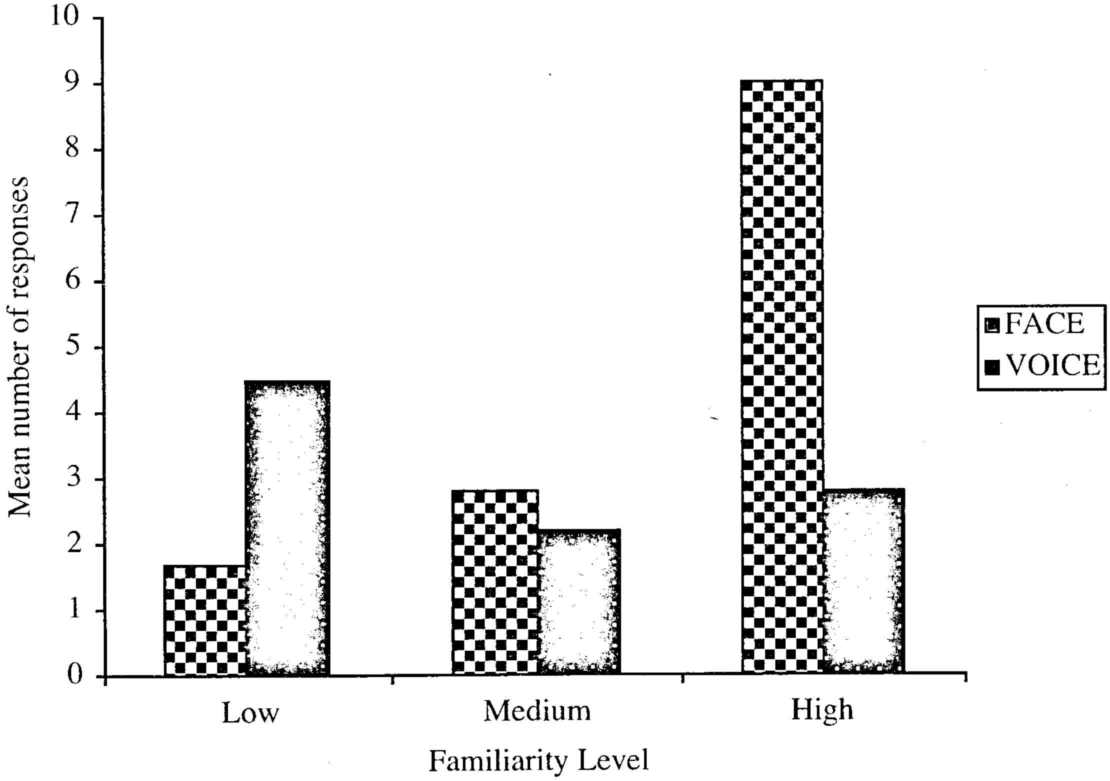
A series of related samples t-test were performed on all of the performance measures for the two modalities. Significantly more famous items were recognized in the face condition than in the voice condition, t (29) = 8.52, $p < 0.05$. False alarms on the other hand were more likely to occur in response to voices than faces, t (29) = 4.69, $p < 0.05$. In terms of the number of occupations and names recalled, faces were associated with better recall of occupations and names than voices, t (29) = 10.19, $p < 0.05$ and t (29) = 9.33, $p < 0.05$, respectively. The same pattern emerged when occupations and names were conditionalized on familiarity ( $p < 0.05$).

### *Signal Detection Measures.*

The ability to discriminate famous items from non-famous ones was significantly better in response to faces than voices. A' measures were 0.91 for faces and 0.69 for voices, respectively. Both were significantly greater than chance ($p < 0.05$). Measures of bias as indicated by B"D scores was 0.25 for faces and 0.13 for voices. Both bias measures were in the conservative direction. However the result from a one-sample t-test showed that both bias measures did not differ from neutral ($p > 0.05$) nor did they differ statistically from each other ($p > 0.05$).

The number of correct responses at each level of familiarity across the two modalities was assessed by a 3x2 repeated measures analysis of variance. This produced a main effect of familiarity, F (2, 58) = 23.49, $p < 0.05$ and a main effect of modality, F (1, 29) = 72.59. The interaction between familiarity and modality was also significant, F (2, 58) = 48.53, $p < 0.05$. Tukey tests showed that there were more low familiarity responses made to voices than faces ($p < 0.05$). At medium and high levels of familiarity, there were significantly more responses to faces than voices ($p < 0.05$). For voices, there were significantly more responses made at low levels of familiarity than medium and high levels of familiarity, which did not differ from each other. For faces, this pattern of responding was reversed (see figure 3.3.a).

Figure 3.3.a. Mean number of responses at each level of familiarity for faces and voices.

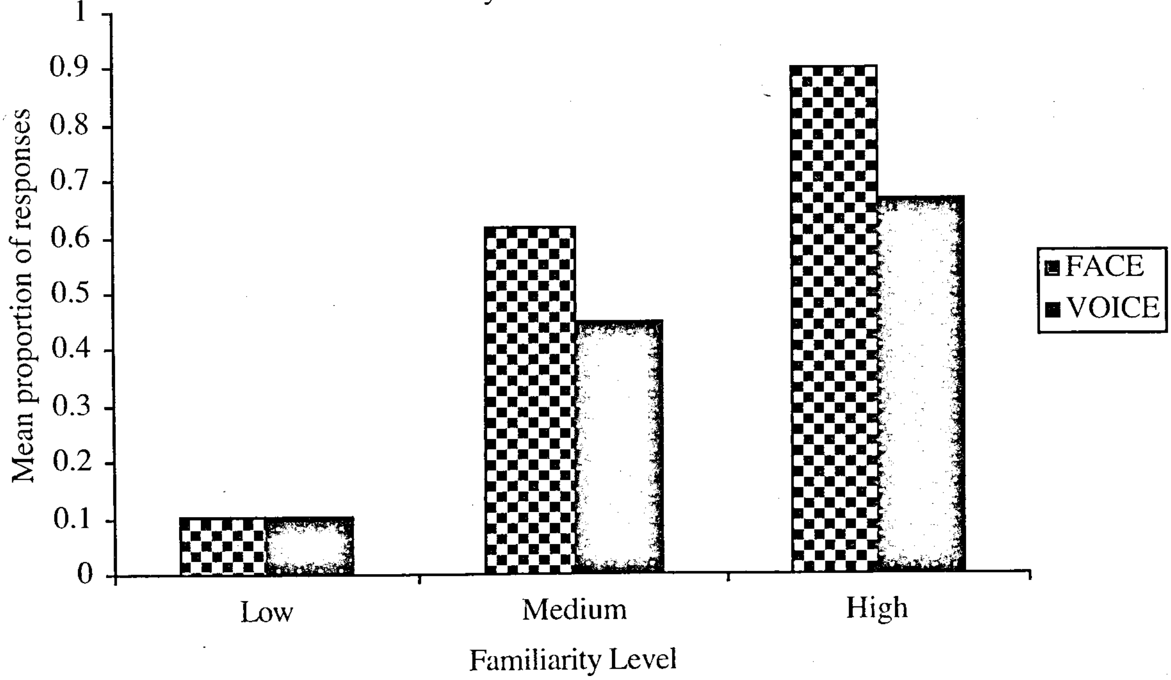


*Retrieval of person-specific identity information.*

*Occupational Recall*

Retrieval of occupational information once conditionalized on familiarity was examined by a 3x2 repeated measures analysis of variance. This produced a main effect of familiarity, F (2, 58) = 99.57, $p < 0.05$ and a main effect of modality, F (1, 29) = 6.70, $p < 0.05$. The familiarity by modality interaction was also significant, F (2, 58) = 3.92, $p < 0.05$. Tukey tests revealed that there were significantly more occupations recalled at medium than low levels of familiarity and also more occupations recalled at high than medium levels of familiarity ($p < 0.05$). This pattern is the same for both faces and voices. Comparisons across the two modalities showed that whilst performance was the same at low and medium levels of familiarity, at high levels of familiarity there were significantly more occupations retrieved to faces than voices (see figure 3.3.b). An analysis by items also produced a main effect of modality F (1, 31) = 29.04, $p < 0.05$.

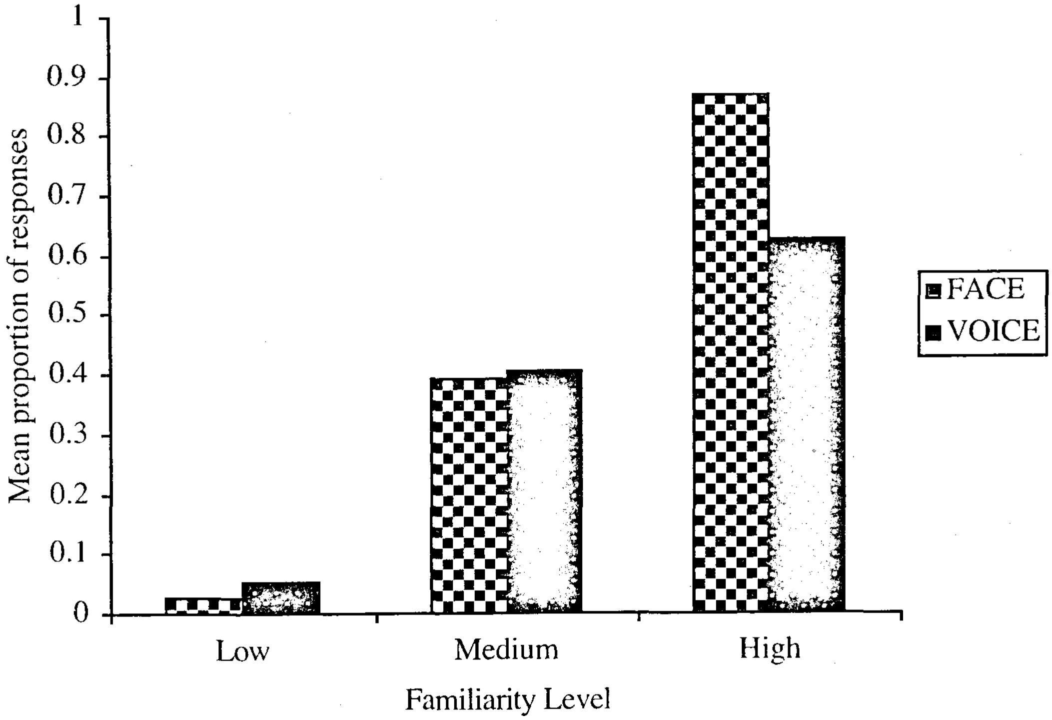
Figure 3.3.b. Mean proportion of occupations recalled at each level of familiarity for faces and voices.



*Name Recall*

A 3x2 repeated measures analysis of variance was also administered on the proportion of names conditionalized on familiarity. This produced a main effect of familiarity F (2, 58) = 116.05, $p < 0.05$. Surprisingly a main effect of modality was not found, F (1, 29) = 2.09, p = 0.159. However, the familiarity by modality interaction was significant, F (2, 58) = 6.35, $p < 0.05$. Tukey tests revealed a similar pattern to the occupational data. For both modalities, the proportion of names recalled increased significantly with increases in familiarity ($p < 0.05$). The only difference between the two modalities was at high levels of familiarity, where there were significantly more names retrieved to faces than voices ($p < 0.05$). This pattern of performance is displayed in figure 3.3.c.

Figure 3.3.c. Mean proportion of names recalled at each level of familiarity for faces and voices.



## Discussion.

Experiment 1c attempted to increase the opportunity for calibration by designing the experiment so that participants alternated between faces and voices on each trial. Previous research has shown that when two tasks are presented in an interleaved format, participants become better calibrated in terms of their memories and their confidence ratings (Hollins and Perfect, 1997) . However, this was not the case for familiarity judgments to faces and voices. Even under these circumstances, retrieval of occupational information was not the same for the two modalities even when familiarity was equated. Furthermore, analysis of the occupational recall data showed that faces given a 'high' familiarity rating were significantly more likely to be associated with occupational details than a voice that was given a 'high' familiarity rating. Similar observations were made with the name recall data. This indicates that there are differences between voices and faces at the level which familiarity decisions are taken.

One area of concern across the three experiments was that overall voice recognition performance was quite low. The hit rates in the voice conditions across the three experiments when converted in the percentages were, 56%, 58% and 59% in Experiments 1a, b and c, respectively. In these experiments, the duration of each clip was approximately 7 seconds. When this is considered in the context that 2 seconds of speech can produce hit rates of 68% (Schweinberger, 2001), then the low hit level in the voice condition becomes even more apparent.

During the de-briefing process, many participants noted that they found it particularly hard to discriminate between the famous American voices and the non-famous American voices. They would often claim that these two categories of voices sounded all the same. This was investigated further by calculating A' values for each participant for American faces, American voices, British faces and British voices for each experiment (1a, 1b and 1c) separately. It was anticipated that participants' difficulty with the American voices should be shown in the form of an interaction effect. Specifically, it was expected that face recognition should be better than voice recognition, irrespective of nationality. However, if American voices lead to poor discrimination between famous and non-famous items than British voices, then lower A' values should be more strongly associated with American voices than British voices.

Participants' A' values for the four measures were entered into a 2 (modality: faces and voices) by (nationality: American and British) mixed ANOVA for Experiment 1a. There was a main effect of modality F (1, 48) = 33.5, $p < 0.05$. The main effect of nationality was not significant. Contrary to predictions, the modality by interaction failed to reach significance ($p = 0.14$). This analysis simply showed that face recognition was better in response to faces than voices, irrespective of stimulus nationality.

For experiment 1b, the repeated measures ANOVA produced a main effect of modality, F (1, 29) = 22.4, $p < 0.05$, but no effect of nationality. The interaction was significant, F (1, 29) = 7.23, $p < 0.05$. Although there was a trend for participants to have higher A' values to British voices (0.61) than American voices (0.54), the difference between the two modalities was not significant ($p > 0.05$).

For experiment 1c, the repeated measures ANOVA produced a main effect of modality, F (1, 29) = 49.4, $p < 0.05$, no main effect of nationality, however the interaction was significant, F (1, 29) = 6.92, $p < 0.05$. Again there was a trend for participants to produce higher A' values for British voices (0.65) than American voices (0.54), planned comparisons t-tests revealed that this difference was statistically significant ($p < 0.05$).

This series of analysis provides a tentative suggestion that participants (especially in Experiment 1c) found it hard to discriminate famous from non-famous voices when they were spoken with an American accent. This particular outcome may represent what is commonly referred to as the other-race effect in the face recognition literature (Malpass and Kravitz, 1969). Experimental studies with unfamiliar faces has shown that participants tend to be more accurate at recognizing faces of their own race than they are at recognizing faces of a different race. The findings from this study suggest that this other-race effect may generalize into the auditory domain (Goggin, Thompson, Strube and Simental, 1991). What makes the results from the above analysis interesting is that participants show a marked reduction in sensitivity for voices spoken in the same language. Therefore differences between American and British voices can not be attributed to language familiarity differences, per se. Instead the way in which words are pronounced may be more salient with British voices than American ones. Furthermore, these cues in British voices can be used to extract social characteristics , such as regional dialects and social status (Sapir, 1927) which may account for the advantage of British over American voices.

On the basis of these findings Experiment 1a and 1c were repeated with a new selection of stimuli. In these experiments, all items (famous and non-famous, faces and voices) were of British nationality in order to match the nationality of the participants.

## Method.

### *Experiment 1d.*

### Participants.

Forty participants, the majority of whom were undergraduate students at the University of Essex, took part in the experiment. The age range of the participants was 18 to 40 years, with a mean age of 24.2 years (30 female and 10 male). All participants were British and had normal to corrected-to-normal vision and normal hearing. All participants were paid £5.00 for their participation.

### Apparatus.

Same as previous experiments.

### Stimuli.

The selection of stimuli used in this experiment differed from the ones used in Experiments 1a, 1b and 1c. Specifically, an additional 45 hours of television recordings were taken from 2001 – 2002. The recordings covered interview programs, chat shows and documentaries. These recordings generated an extra 50 British celebrities from various backgrounds in the media, sport, television, film, comedy, music and politics. An additional 40 non-famous British people were selected from the recordings. The new collection of famous people were rated in the same format as described in Part 1 of Chapter 2.

A separate group of 12 participants were presented with each celebrity's name and asked to rate their familiarity with each person's name, face and voice on a familiarity rating scale ranging from 0 (Unfamiliar) – 3 (High familiarity). The rating study produced means of 2.85 for names (SD = 0.60), 2.74 for faces (SD = 0.68) and 2.17 (SD = 0.58) for voices, respectively.

Faces and voices were prepared in the same way as outlined in Part 2a and 2b in Chapter 2. For the face condition 14 famous faces were eliminated due to poor picture quality. For the voice condition, 1 famous person was correctly identified and subsequently discarded. At this stage, there were 35 famous items 40 non-famous items remaining. From the original stimulus set created for experiments 1a, 1b and 1c the 15 British celebrities with the highest voice familiarity ratings were added to this new set. An additional 10 non-famous British people were also selected from experiments 1a, 1b and 1c.

The 100 (50 Famous and 50 Non-Famous) items were transcribed into a written format and presented to a new group of 10 participants. The transcripts were presented in the same fixed random order as used in experiment 1d and 1e. Participants were required to indicate whether they associated each statement with a Famous (F) or Non-famous (N) person. This stage of the assessment yielded means of 44% and 55% for famous and non-famous statements, respectively. Consistent with the previous assessment done in Part 2b, the difference between the means was not significant, $t(9) = 2.03$, $p > 0.05$.

Participants in this pilot study were also administered with the second stage of this assessment. As before, participants were asked if they could associate the written text to a particular celebrity. In 25 out of 1000 opportunities (10 participants x 100 written formats), participants offered a guess of the target person. However, none of these guesses were correct.

After each item was screened and corrected for visual and auditory cues, the familiarity ratings were recalculated for the 50 famous items. This generated a familiarity score of 2.84 (SD = 0.62) for names, 2.68 (SD = 0.65) for faces and 2.20 (SD = 0.60) for voices, respectively. In experiments 1d, and 1e, there were a total of 48 famous items used and 48 non-famous items used. The remaining 2 famous and 2 non-famous items were allocated to the practice trials. Appendix 3 lists the names of the celebrities used in this experiment. The average duration for each item was approximately 8 seconds.

### Design.

This experiment used a between subjects design. Twenty participants were randomly allocated to the face condition and twenty participants were randomly allocated to the voice condition. The dependent measures in this experiment were the same as in the previous experiments.

### Procedure.

The procedure for this experiment was essentially the same as in Experiment 1a. Participants in the face condition were presented with the person's face without an auditory output. Participants in the voice condition were presented with the person's voice, which was played through a set of Sennheiser headphones. The experimental trials were arranged in a fixed random order. The 96 experimental trials were presented within two blocks of 48 trials. Each block had an equal number of famous (24) and non-famous (24) items. Presentation of each block was counterbalanced across participants.

As in previous experiments participants were required to give a familiarity rating (0 – 3) for each face or voice that was presented to them. For each face/voice that was given a 1,2 or 3 familiarity rating, participants were required to retrieve occupational details about the person and to name them.

## Results.

Table 3.4.a provides the overall recognition performance across the two modalities. The mean number of hits is out of a total of 48 and the false alarms are out of a total of 48 for each condition.

### *Overall recognition performance.*

Table 3.4.a. Means and standard deviations (in parentheses) for the face and voice conditions.

| | Faces | Voices |
|---|---|---|
| Hits | 41.8<br>(4.4) | 29.4<br>(4.5) |
| False alarms | 5.45<br>(5.2) | 11.8<br>(6.7) |
| Number of occupations correct | 33.5<br>(7.9) | 10.8<br>(5.7) |
| Number of names correct | 30.3<br>(8.2) | 8.5<br>(3.7) |
| Percentage of occupations recalled conditionalized on familiarity | 80.0<br>(17.4) | 37.0<br>(21.4) |
| Percentage of names recalled conditionalized on familiarity | 72.6<br>(18.4) | 28.8<br>(12.5) |

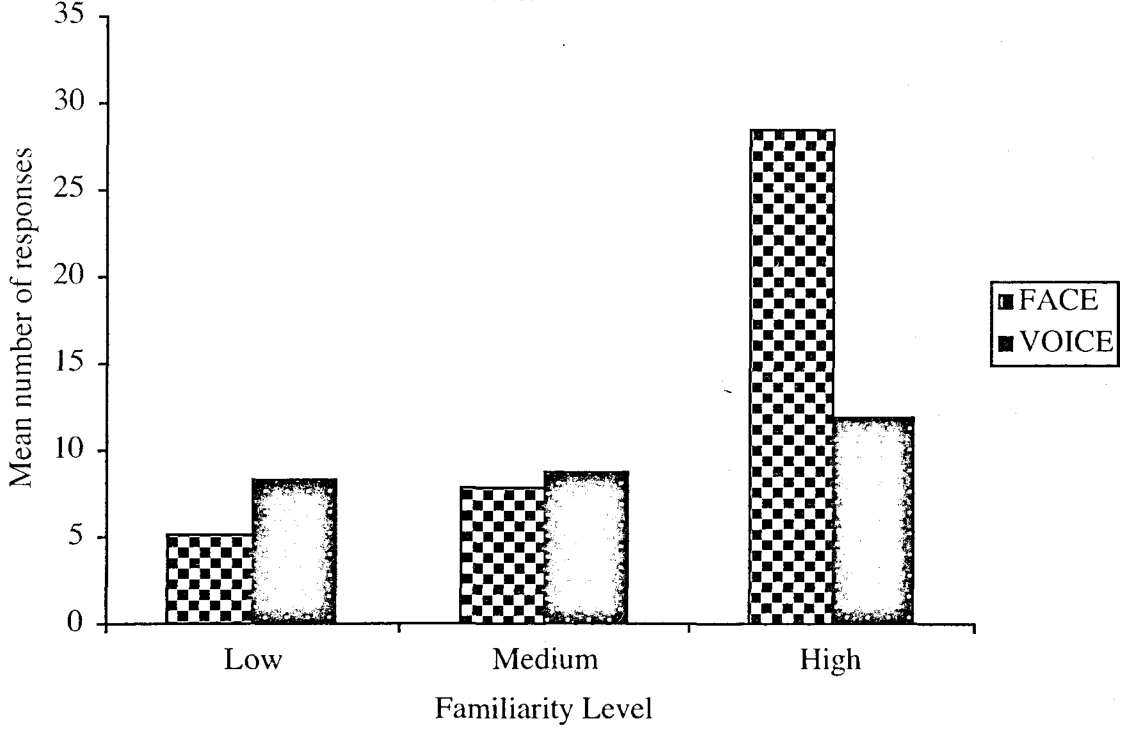
An independent samples t-test showed that there were significantly more hits made to faces than voices, t (38) = 8.70, $p < 0.05$. Furthermore, there were significantly fewer false alarms made to faces than voices, $t\ (38) = 3.32$, $p < 0.05$. Occupational and name retrieval performance showed that there were significantly more occupations and names retrieved to faces than voices ($p < 0.05$). Superiority for faces over voices was maintained even when occupational and name retrieval were examined as a function of overall familiarity ($p < 0.05$).

### *Signal Detection Measures.*

The ability to discriminate between famous and non-famous items was significantly better in response to faces ($A' = 0.93$) than voices (0.76), $p < 0.05$. Both measures differed significantly from chance. Measures of bias produced values that were in the conservative direction for both modalities, $B''D = 0.16$ for faces and $B''D = 0.33$ for voices, respectively. Bias differed significantly from neutral for the voice condition ($p < 0.05$) but not for the face condition ($p > 0.05$). The degree of bias between the two modalities did not differ significantly from each other ($p > 0.05$).

The number of correct familiarity decisions made at each level of familiarity across the two modalities was analyzed by a 3x2 mixed analysis of variance. This produced a main effect of familiarity $F\ (2, 76) = 39.94$, $p < 0.05$ and a main effect of modality $F\ (1, 38) = 58.55$, $p < 0.05$. The familiarity by modality interaction was also significant, $F\ (2, 76) = 26.10$, $p < 0.05$. The nature of this interaction was analyzed by a series of Tukey tests which showed that there were more responses made at low levels of familiarity to voices than faces ($p < 0.05$). For high levels of familiarity, there were more responses to faces than voices ($p < 0.05$). For voices, the number of responses made at each level of familiarity did not differ. This was not true for the face condition, which produced more responses at high levels of familiarity than at medium and low levels of familiarity ($p < 0.05$). This pattern of performance is displayed in figure 3.4.a.

Figure 3.4.a. Mean number of responses at each level of familiarity for faces and voices.

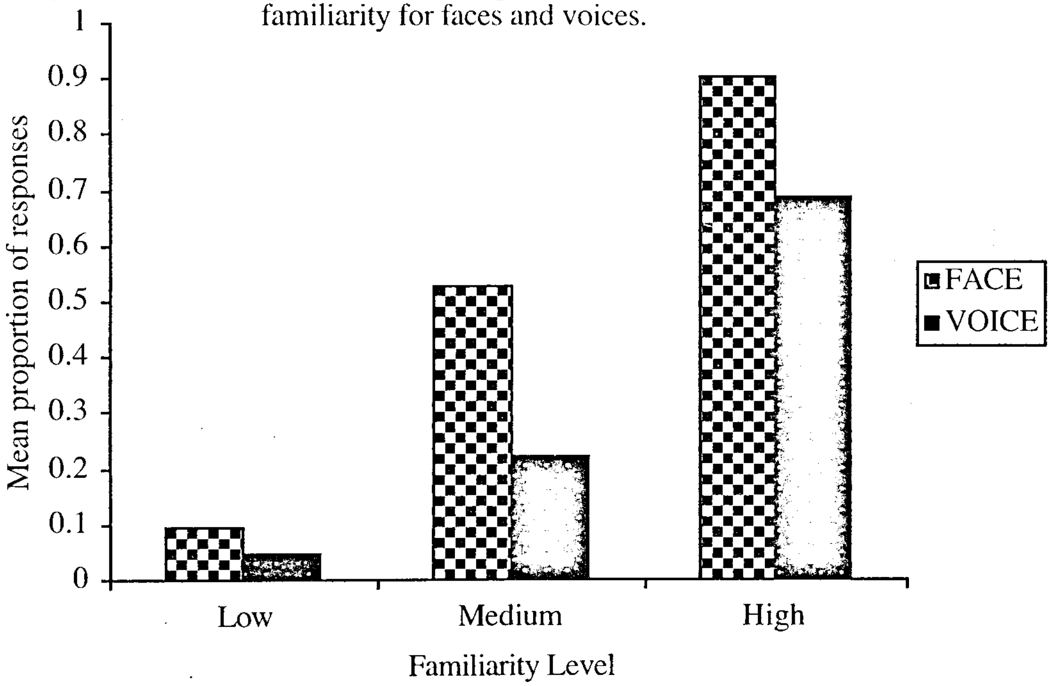


*Retrieval of person-specific identity information.*

*Occupational recall.*

The proportion of occupations recalled conditionalized on each level of familiarity was analyzed by a 3x2 mixed analysis of variance. This produced a main effect of familiarity $F(2, 76) = 216.88$, $p < 0.05$, a main effect of modality $F(1, 38) = 21.18$, $p < 0.05$ and a significant familiarity by modality interaction $F(2, 76) = 6.71$, $p < 0.05$. Tukey tests showed that for both modalities, the proportion of occupations recalled increased significantly with each increase of familiarity ($p < 0.05$). Comparison of occupational retrieval across modalities shows that there were significantly more occupations retrieved to faces than voices at medium and high levels of familiarity ($p < 0.05$). There was not a significant difference between the two modalities at low levels of familiarity (see figure 3.4.b). There was also a main effect of modality by items, F (1, 94) = 68.51, $p < 0.05$.

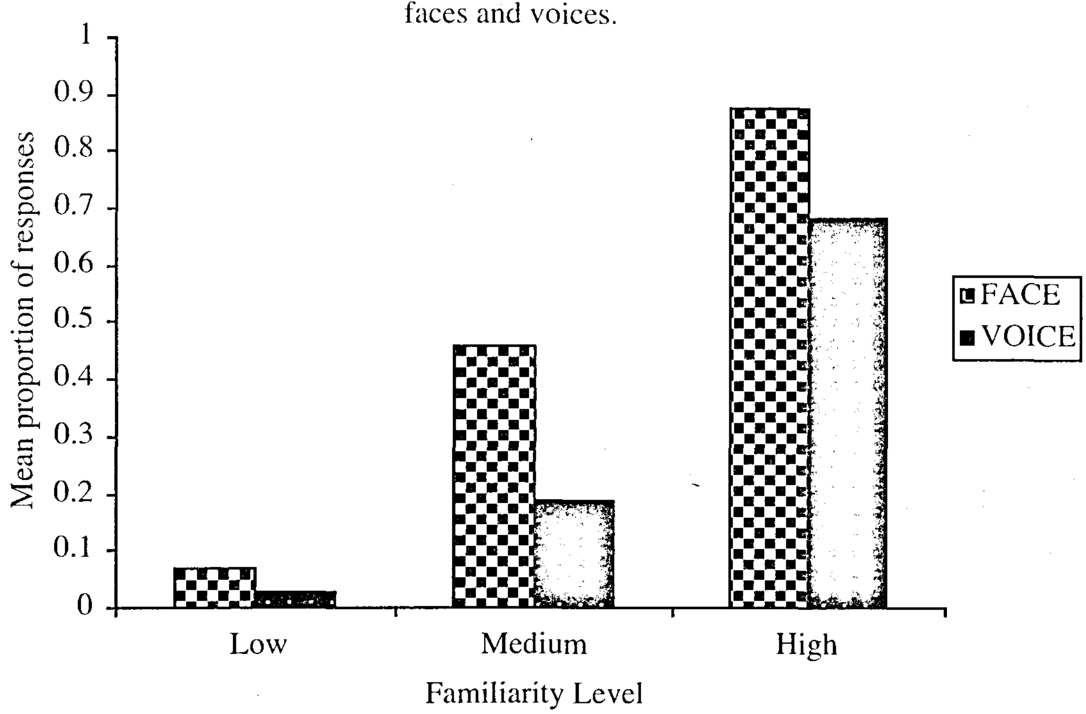
Figure 3.4.b: Mean proportion of occupations recalled at each level of familiarity for faces and voices.



*Name recall.*

The proportion of names recalled as conditionalized on familiarity was analyzed by a 3x2 mixed analysis of variance. This produced a main effect of familiarity, F (2, 76) = 211.41, $p < 0.05$, a main effect of modality, F (1, 38) = 15.34, $p < 0.05$ and a significant interaction, F (2, 76) = 5.10, $p < 0.05$. Tukey tests showed that for both modalities, the proportion of names retrieved increased as familiarity increased ($p < 0.05$). Across modalities, there were significantly more names retrieved to faces than voices at medium and high levels of familiarity ($p < 0.05$). See figure 3.4.c.

Figure 3.4.c. Mean proportion of names recalled at each level of familiarity for faces and voices.



**Discussion.**

The findings from this experiment extend the results from Experiment 1a. The aim of repeating experiment 1a was to improve recognition in the voice condition by using a selection of voices that controlled for potential nationality effects. The mean A' measure for the voice conditions calculated across experiments 1a, 1b and 1c was 0.703. When this value was used in a one sample t-test, it showed that voice recognition in experiment 1d with the new stimuli (0.76) was significantly better, t (19) = 2.67, $p < 0.05$. In other words, the ability to discriminate famous from non-famous voices was better in Experiment 1d than in the previous experiments. Despite this, participants are still better at retrieving occupational and name information in response to seeing a familiar face than hearing a familiar voice.

The final experiment in this chapter investigates whether alternating between faces and voices would encourage participants to use the same criteria for their familiarity decisions. Therefore, Experiment 1c was repeated with this new set of stimuli in a within-subjects experimental design.

## Method.

### *Experiment 1e.*

**Participants.**

Twenty-two participants, the majority of whom were undergraduate students at the University of Essex, took part in the experiment. The age range of the participants was 18 to 48 years, with a mean age of 22.4 years (18 female and 4 male). All participants were British and had normal to corrected-to-normal vision and normal hearing. All participants were paid £5.00 for their participation.

**Apparatus.**

Same as previous experiments.

**Stimuli.**

The stimuli used were the same as in Experiment 1d.

**Design.**

This experiment used a within subjects design with all participants exposed to faces and voices. The dependent measures in this experiment were the same as in the previous experiments.

**Procedure.**

The procedure for this experiment was essentially the same as in Experiment 1c. Each famous and each non-famous item was presented as a face to half of the participants and as a voice to the other half. The 22 participants alternated between faces and voices on every trial. The experimental trials were arranged in a fixed random order.

The 96 experimental trials consisted of 48 famous clips and 48 non-famous clips. The 96 experimental trials were presented within two blocks of 48 trials. Each block had an equal number of famous (24) and non-famous (24) items. Presentation of each block was counterbalanced across participants. As participants were exposed to both face and voice modality conditions, the number of famous faces and voices used in this experiment was 24 for each modality. The number of non-famous faces and voices was 24 for each modality.

## Results.

Table 3.5.a provides the overall recognition performance across the two modalities. The mean number of hits is out of a total of 24 and the false alarms are out of a total of 24 for each condition.

### *Overall recognition performance.*

Table 3.5.a. Means and standard deviations (in parentheses) for the face and voice conditions

| | Faces | Voices |
|---|---|---|
| Hits | 22.4<br>(1.6) | 17.3<br>(4.0) |
| False alarms | 4.7<br>(2.9) | 8.4<br>(4.3) |
| Number of occupations correct | 18.1<br>(3.4) | 6.0<br>(3.2) |
| Number of names correct | 15.6<br>(4.3) | 5.0<br>(3.2) |
| Percentage of occupations recalled conditionalized on familiarity | 81.0<br>(12.0) | 34.0<br>(14.6) |
| Percentage of names recalled conditionalized on familiarity | 69.0<br>(15.9) | 28.3<br>(15.0) |

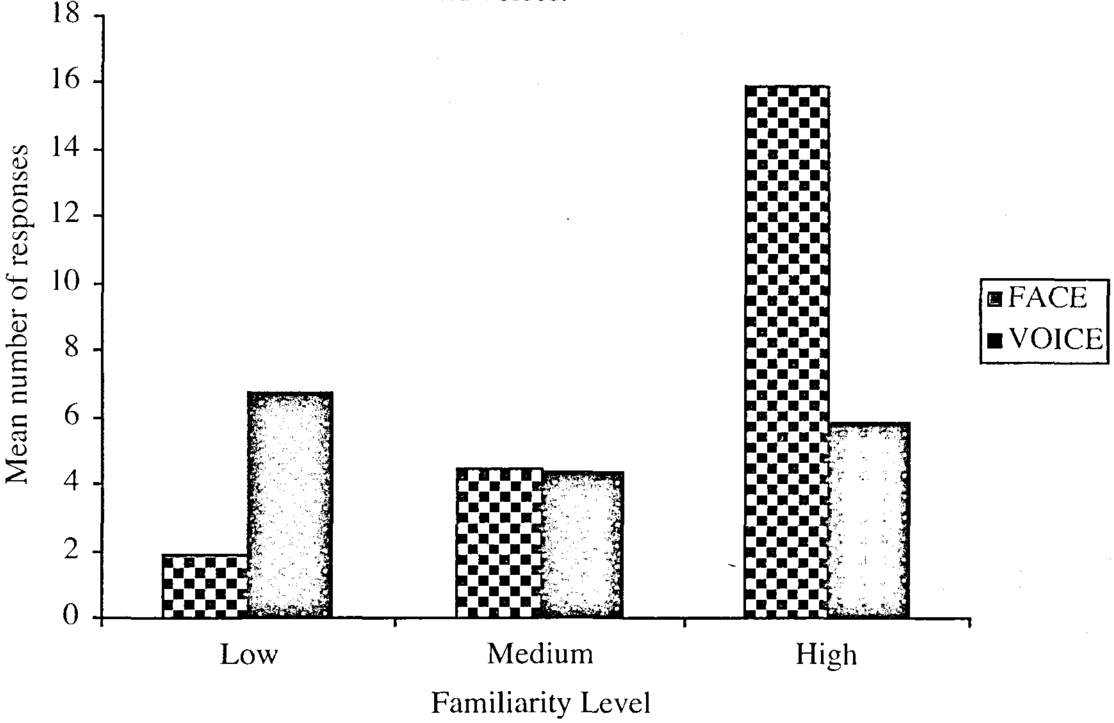
A series of paired samples t-test revealed significantly more hits made in response to faces than voices, t (21) = 6.84, $p < 0.05$ and significantly fewer false alarms made to faces than voices, t (21) = 3.48, $p < 0.05$. There were significantly more occupations and significantly more names recalled to faces than voices, ($p < 0.05$) even when conditionalized on familiarity ($p < 0.05$).

### *Signal Detection Measures.*

The ability to discriminate famous items from non-famous ones was significantly better in the face condition (0.93) than in the voice condition (0.77), $p < 0.05$. Measures of discriminability were significantly above chance in both conditions ($p < 0.05$). Measures of bias was in the liberal direction for faces (-0.52) and voices (-0.13). Whilst the bias measure in the face condition significantly differed from neutral ($p < 0.05$), the degree of bias in the voice condition did not differ from neutral ($p > 0.05$). Furthermore, responses in the face condition was statistically more liberal than in the voice condition ($p < 0.05$).

The number of correct responses made at each level of familiarity across the two conditions was analysed by a 3x2 repeated measures analysis of variance. This produced a main effect of familiarity, $F(2, 42) = 20.08$, $p < 0.05$, a main effect of modality, $F(1, 21) = 59.19$, $p < 0.05$ and a significant familiarity by modality interaction, $F(2, 42) = 100.69$, $p < 0.05$. Tukey tests showed that there were more responses made to voices than faces at low levels of familiarity ($p < 0.05$). This pattern of responding was reversed at high levels of familiarity ($p < 0.05$). Whilst the number of correct responses at each level of familiarity did not differ in the voice condition, the face condition produced more responses at high levels of familiarity than at low and medium levels of familiarity ($p < 0.05$). See figure 3.5.a.

Figure 3.5.a. Mean number of responses at each level of familiarity for faces and voices.

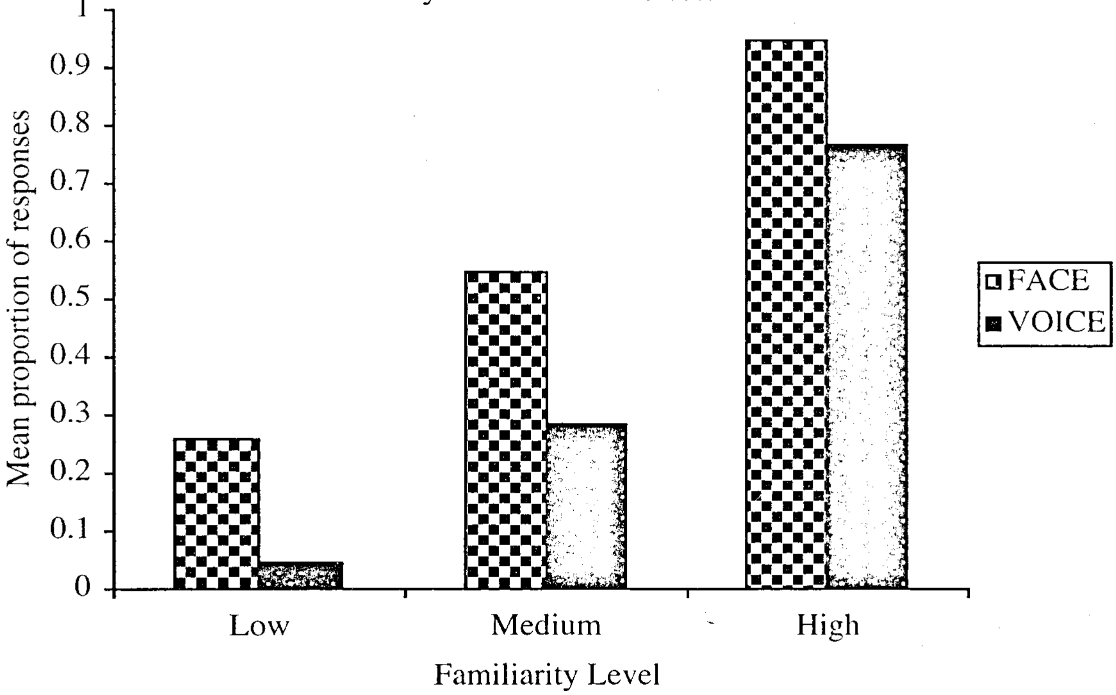


### *Retrieval of person-specific identity information.*

#### *Occupational recall.*

Participants' ability to retrieve occupational details to faces and voices that they previously found familiar was assessed by a 3x2 repeated measures analysis of variance. This produced a main effect of familiarity F (2, 42) = 112.96, $p < 0.05$ and a main effect of modality, F (1, 21) = 44.72, $p < 0.05$. The interaction was not significant. This analysis shows once again that there were more occupations retrieved to faces than voices even when levels of familiarity were equated. See figure 3.5.b. There was also a main effect of modality by items, (1, 47) = 71.14, $p < 0.05$.

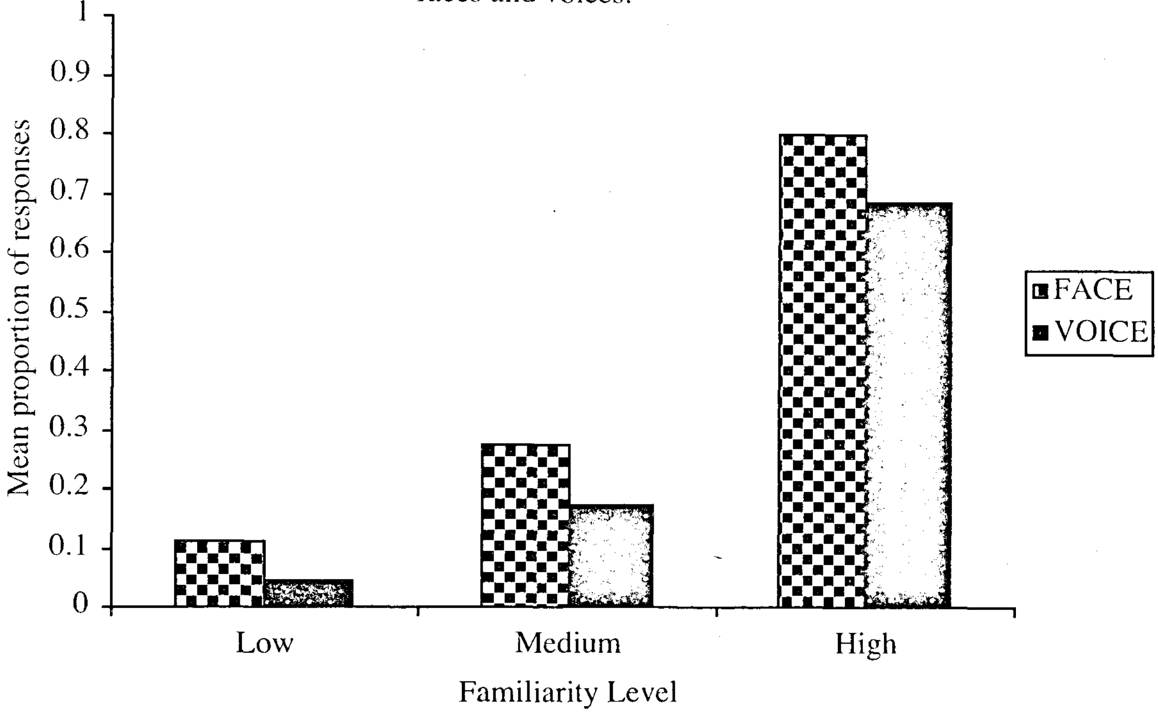
Figure 3.5.b. Mean proportion of occupations recalled at each level of familiarity for faces and voices.



*Name recall.*

Participants' ability to retrieve a name to a face or voice that they previously found familiar produced similar results to the occupational recall data. A 3x2 repeated measures analysis of variance produced a main effect of familiarity, F (2, 42) = 95.62, $p < 0.05$ and a main effect of modality, F (1, 21) = 16.35, $p < 0.05$. The familiarity by modality interaction was not significant. This analysis shows there were more names associated with faces than with voices. See figure 3.5.c.

Figure 3.5.c. Mean proportion of names recalled at each level of familiarity for faces and voices.



## Discussion.

Experiment 1c used a within-subjects design whereby the same participants were required to make familiarity judgements to both faces and voices with the two modalities alternating on every trial. Experiment 1c failed to remove the familiar-only bias for voices, even when the opportunity for calibration was heightened. Experiment 1e largely extended these results with a new selection of voices. The A' measure for voices in experiment 1e was significantly higher (0.77) than in experiments 1a, 1b and 1c, t (21) = 3.33, $p < 0.05$. This new selection of voices therefore made it easier for participants to discriminate between famous and non-famous items. However, even under these circumstances, occupational and name retrieval remained particularly elusive for voices relative to faces.

## General Discussion.

The five experiments reported in this chapter looked at participants' abilities to recognize famous faces and voices. The key area of interest in this chapter was to examine participants' abilities to retrieve biographical information, such as occupational details, to faces and voices that had previously been rated as familiar. The inability to retrieve contextual information, such as a person's occupation is characteristic of the person recognition error known as the 'familiar only' experience (Young et al, 1985).

Studies of person recognition failures have been used to shed light on the functional components of contemporary person recognition models (Bruce and Young, 1986, Burton, Bruce and Johnston, 1990), with specific emphasis on elucidating the face recognition route (Hanley and Cowell, 1988). With the exception of Hanley, Smith and Hadfield (1998), very few studies have attempted to investigate person recognition errors across different modalities within the same study. This is rather surprising, given that the functional validity of such models can be strengthened if they can adequately accommodate performance measures for both faces and voices.

Hanley et al's (1998) study demonstrated that overall recognition for faces was significantly better than voices – a finding which was replicated across the five experiments reported here. Using a set of new stimuli, the average hit rate for the faces across the five experiments was 0.84, a value which is similar to Hanley et al's (1998) average hit rate of 0.92 across their two experiments. The average hit rate in the voice condition for Experiments 1a, 1b and 1c was 0.57, which is somewhat lower than the hit rate achieved in Hanley et al's voice condition (0.70). A selection of new items used in experiment 1d and 1e raised the average hit rate of 0.67, which is at a comparable level to Hanley et al's (1998) and Schweinberger's (2001) value.

In Experiments 1b, 1c, 1d and 1e, false alarm rates were significantly more frequent in response to voices than faces. The only exception was experiment 1a, where the false alarm rates did not differ between the two modalities (see table 3.1.a). Even with a new selection of stimuli, the average false alarm rates observed across the five experiments was comparable to previous values obtained by Hanley et al (1998). Hanley et al's false alarm rates for their face condition was 0.19 compared to 0.17 across the five experiments in this study. For the voice condition, Hanley et al obtained an average false alarm value of 0.31, which is exactly the same value as the average false alarm rate calculated across the five experiments reported here. This shows that the hit and false alarm rates observed by Hanley et al (1998) can be easily extended to a new set of stimuli.

Furthermore, the superiority for face recognition over voice recognition with familiar faces and voices parallels previous work with newly learnt faces and voices (Yarmey, Yarmey and Yarmey, 1994). In this particular study, participants engaged in a conversation with a stranger for 15 seconds. After a period of 5 minutes, participants were required to identify the person either visually by selecting a photograph or by listening to a recording of the person's voice. The visual condition produced an average hit rate of 52% compared to 19% for the voice recognition. Whilst these recognition rates are both considerably lower than those obtained from familiar faces and voices, the difference between the two modalities displayed with newly learnt faces and voices is similar to familiar items.

However, Hanley et al's (1998) findings did not merely demonstrate that familiar faces are easier to recognize than voices. When we see a face that we find familiar, we are also able to retrieve person specific identity information about the person, such as the person's occupation. Previous research by Hay, Young and Ellis (1991), has shown that participants could retrieve appropriate person specific information on approximately 94% of occasions. In Hanley et al's study, the ability to retrieve occupational details to a face that was subsequently found familiar was also quite successful, with participants being able to do so on 89% of occasions across their two experiments. However, this pattern of success was not observed with participants in the voice condition. Across their two experiments, participants in the voice condition averaged a mere 58% retrieval rate of occupations to voices that they previously deemed as familiar.

Experiments 1a and 1d, using a between-subjects design replicated these initial findings observed by Hanley et al (1998). In experiment 1a, the proportion of occupations retrieved to familiar faces was 0.82 compared to 0.50 for voices. Even when the ability to discriminate famous voices from non-famous ones was increased by using a better set of stimuli in experiment 1d, performance was significantly worse in the voice condition. In experiment 1d the occupational retrieval rate was 0.80 for faces and 0.37 for voices, respectively.

The occupational retrieval rates across all the experiments in this chapter are somewhat lower than those observed in Hanley et al's (1998) study. This to some extent may be accounted for by the strict marking criterion employed in assessing occupational recall (see page 75). Whilst other researchers have assessed occupational recall by adopting a hierarchical format (see Haslam, Cook and Coltheart, 2001), for present purposes the current results indicate when non-generic occupational information is required, participants are much worse at retrieving such information in response to hearing the voice of a celebrity compared to seeing their face.

In experiment 1a, participants retrieved significantly more occupations to faces than voices, even when levels of familiarity were held constant (see figure 3.1.b). Similar results were obtained in experiment 1d, which also utilised a between-subjects design (see figure 3.4.b). However, in this experiment, occupational recall was quantified by a significant familiarity by modality interaction. Post-hoc tests indicated that the greatest difference between the two modalities was at high levels of familiarity. The data displayed in figures 3.1.b and 3.4.b indicate that there are differences between faces and voices in the level at which familiarity decisions are made. One explanation of these data is that the decision a face is deemed 'highly' familiar is not based on the same criteria as a voice that is deemed 'highly' familiar. It seems that 'very familiar' does not mean the same thing in the face condition as in the voice condition.

As such the data are very similar as those reported by Hanley et al (1998). However, as outlined in the introduction section of this chapter, the between-subjects design employed by Hanley et al (1998) may have concealed criteria differences between participants in the two different modality conditions. Experiments 1b, 1c and 1e aimed to test the possibility that providing participants with their own context of familiarity would help participants to become better calibrated in terms of their familiarity rating for voices and faces. This idea was tested by administering both face and voice stimuli to the same participants, by utilising a within-subjects design.

This subtle, yet effective change in experimental design has been most notably demonstrated in the confidence-accuracy literature (C-A), as reviewed in section 3.4 of this chapter. The C-A literature consistently shows low to non-significant confidence-accuracy correlations for eyewitness memory, but modest and significant confidence-accuracy correlations for general knowledge (see Perfect, Watson and

Wagstaff, 1993). This has led some authors to argue that the qualitative factors that differentiate eyewitness memory from general knowledge prevent participants from developing an insight into their eyewitness memory abilities, thus leading non-significant C-A correlations. Some qualitative factors that have been proposed, include the emotional demands of eyewitness memory (Perfect et al, 1993) and that eyewitness events tend to be perceptually richer and more complex than general knowledge items (Smith, Kassin and Ellsworth, 1989). Despite these differences, Perfect and Hollins have shown that when the two memory tests are presented in a mixed-question type format, then the confidence-accuracy relationship for eyewitness memory increases and in some cases reaches significance compared to a condition where participants have to complete each test separately. Perfect and Hollins (1997) have argued that by mixing the two tests together, participants can use their confidence judgements for general knowledge items as a context to calibrate themselves for eyewitness memory items.

It was anticipated that by providing participants with their own familiarity judgements for faces in experiments 1b, 1c and 1e, participants could use this context to help them assign their familiarity judgements for voices. In experiment 1b, participants alternated between faces and voices after every fourth trial. Under these circumstances, participants still continued to produce fewer occupations to voices (0.46) than faces (0.84). When occupational performance was investigated at each level of familiarity, participants showed good calibration within each modality across the varying levels of familiarity in that high familiar faces led to better recall of occupations than faces deemed lower in familiarity (see figure 3.2.b). Crucially, figure 3.2.b also shows that faces tended to generate significantly more occupations than voices at each level of familiarity. Therefore there is no evidence that providing participants with their own familiarity context can eliminate differences in retrieval of biographical information between faces and voices deemed familiar. This means that the difference between faces and voices is not one of calibration.

Using the same stimuli, experiment 1c aimed to increase the contextual effect of familiarity by presenting faces and voices on each trial. Even under these circumstances, there were more occupations recalled to faces deemed familiar (0.74) than voices deemed familiar (0.35). Occupational performance at each level of familiarity showed similar performance for the two modalities at low and medium levels of familiarity, but superiority of recall of biographical information for faces over voices at high levels of familiarity (see figure 3.3.b). The results from Experiment 1c show that occupational retrieval remained particularly elusive for voices even when the opportunity for calibration was heightened.

A new selection of stimuli managed to improve the overall ability to discriminate famous from non-famous items in the voice condition (experiment 1e). However, fewer occupations (0.34) were retrieved to voices compared with faces (0.80). Occupational performance at each level of familiarity showed once again that there were significantly more occupations retrieved to familiar faces than voices.

Participants' ability to retrieve a name to a face or voice that they had previously found familiar was also examined in this chapter, however it was not of central relevance as the occupational data. For completeness, the following sections provide a brief overview of the name recall data across the two modalities.

Across the five experiments the average name recall rate was 0.69 for faces and 0.33 for voices. Both of these values are lower than the occupational recall data, which were 0.80 for faces and 0.41 for voices, respectively. The lower performance rates for name recall is therefore consistent with previous research which has shown that people's names are more difficult to retrieve relative to other facts we know about the person (Young et al, 1985a, McWeeny et al, 1985). For faces, name recall performance was similar to the value obtained by Hanley et al (1998) in their face condition (0.72). However, the overall value of 0.33 for the voice condition in this study is somewhat lower than Hanley et al's name recall value of 0.44. When name recall for the two modalities was investigated at each level of familiarity, name recall was significantly better in response to faces than voices. The only occasions when this did not appear to be the case was in experiments 1a and 1c. Here, surprisingly, the main effect of modality was not found (figures 3.1.c and 3.3.c). With experiment 1c however, there was a significant familiarity by modality interaction. This showed that the only difference between faces and voices was at high levels of familiarity. Specifically, more names were retrieved to faces that were deemed 'high' in familiarity than voices that were given 'high' familiarity ratings. This particular pattern of performance in experiment 1c replicates the findings observed with occupational retrieval (see figure 3.3.b).

Throughout this study, the consistent results obtained from the occupational data indicates that there are clear differences in the extent to which occupational information can be retrieved from familiar faces and voices. This suggests that faces are better than voices at eliciting person specific information. Specifically, this outcome is not an artificat of experimental design. In terms of the relationship between familiarity decisions and person identification, the data suggest that familiarity judgments should be separated from person identification. This argument has already been supported by neuropsychological data (de Haan et al, 1989).

There would appear to be at least two different explanations of this pattern of results. The first is that 'high familiar' means exactly the same thing in the voice condition as it does in the face condition. Fewer occupations are recalled in the presence of highly familiar voices than highly familiar faces because biographical information is more strongly associated with faces than it is with voices. That is, for some reason the links in semantic memory between faces and biographical information are particularly strong. The problem with this explanation are the results of Hanley and Turner (2000). They showed that when overall level of performance in the face and voice condition is matched by blurring the faces that biographical information was just as difficult to recall with faces as with voices. If faces were more strongly associated than voices with biographical information in semantic memory then Hanley and Tuner (2000) should have observed superior recall of occupations in the blurred face condition.

The second explanation is that the voices were given higher familiar ratings than they should relative to faces. That is, familiarity level '3' means less familiar when applied to a voice than when applied to a face. (Bearing in mind the results of Hanley and Turner (2000), it is likely that this is a consequence of the lower overall level of performance in the voice condition rather than some intrinsic quality of voices). If this second account was true, we might have expected that a within-subjects design (particularly one such as experiment 1c and 1e in which face and voice trials alternated), would have reduced these criterion differences between voice and face conditions. One would have expected that having to make each voice response immediately after making each face response would have led participants to use similar criteria to judge the level of familiarity in the face and voice condition. From this perspective that recall of biographical information from 'very familiar' voices remains more difficult than recall of biographical information from 'very familiar' faces in a within-subjects design.

Perhaps therefore, strength of familiarity decisions can not be recalibrated as readily as confidence judgements. If so, it would follow that strength of familiarity decisions may be based on something quite different from degree of confidence ratings. This issue was investigated further in chapter 4.

# Chapter 4: Confidence judgments in person recognition memory.

## 4.1: Introduction.

One obvious continuation of the line of research reported in Chapter 3 is to investigate the effects of confidence judgments in face and voice recognition. This chapter therefore compares the relationship between accuracy and confidence in familiar face and voice recognition. Are people over-confident when attempting to distinguish familiar from unfamiliar voices? It also examines the relationship between confidence and familiarity. It may be the case that when participants are assigning their familiarity judgments to faces and voices they are basing their judgments on something other than decisions of confidence. If this is the case, perhaps one would not observe a modality difference in recall of biographical information from faces and voices when participants confidence in their face and voice decisions is equivalent.

## 4.2:Confidence judgments in person recognition memory.

Experiments conducted on recognition memory performance sometimes use a variety of subjective measures, such as confidence measures and remember and know responses. The use of such subjective measures is of both theoretical and practical importance. These subjective measures of memory, in conjunction with standard objective measures such as proportion correct provide an intricate description of the processes involved in recognition memory. The contribution of remember and know responses to the understanding of recognition processes will be confined to a separate section of this chapter and will be followed up in the final study of this thesis.

The cognitive basis of confidence judgments involves the contribution of two components. According to Perfect, Hollins and Hunt (2000) the processes involved are the ease or fluency with which retrieval is achieved and the contribution of preexisting beliefs about the skill under test. Confidence can be measured through the administration of a number of different response scales. These can range from a 5-point scale with 1 denoting not at all confident through to 5 denoting very confident (Perfect, Hollins and Hunt, 2000). Other researchers opt for the percentage scale to capture levels of confidence. For example, Olsson, Juslin and Winman's (1998) scale ranged from 0% (zero probability that the identified person is the culprit) to 100% (certain that the identified person is the culprit). Alternatively, measures of confidence can be captured by two simple categories; sure and unsure as used by Rajaram, Hamilton and Bolton (2002) and Gardiner (2001). The minimization of the confidence categories in these studies was done to see whether confidence can be distinguished from recollective experience as measured by remembering and knowing.

The assessment of confidence on performance accuracy is sometimes taken in the form of Pearson's correlation co-efficients, with values approaching 1.0 indicating perfect confidence to accuracy correspondence. In other words, a participant that reports the highest level of confidence on a given test item is likely to obtain the highest score correct than somebody whose subjective confidence is low. The level of correspondence between the subjective (confidence) and objective (proportion correct) measures is used to calculate the confidence-accuracy (CA) correlation. When studies use binary confidence judgments, such as the one reported by Rajaram et al (2002), then the proportion of correct scores for the 'sure' category are compared with the proportion of correct responses in the 'unsure' category. The difference between the two categories can be measured through the use of Analysis of Variance.

The reason to pursue confidence judgments in the assessment of participants' abilities to recognize familiar faces and voices was primarily driven by the research conducted in the applied forensic field. A general review of the literature showed that confidence judgments are a widely accepted form of protocol used by law enforcement agencies. A nationwide survey of police officers in the United States showed that 86% of police officers use confidence judgments to question witnesses after identification procedures (Wogalter, Malpass and Burger, 1993, cited in Olsson, 2000). These practical applications of confidence judgments raise the question how these judgments are interpreted in other legal contexts.

According to Smith, Kassin and Ellsworth (1989), 56% of jurors give strong predictive weight to confidence judgments, believing that eyewitness confidence and accuracy are positively related (statistic taken from Brigham and Bothwell, 1983, cited by Smith et al, 1989). Other legal professionals hold similar perceptions, with 73% of police officers, 75% of prosecuting attorneys and 40% of defence attorneys viewing eyewitness confidence being positively related to accuracy (Brigham and Wolfskiel, 1983, cited by Smith et al, 1989). Outside of the legal context, 76% of Undergraduates believe confidence to be a good predictor of accuracy (Deffenbacher and Loftus, 1982, cited by Smith et al, 1989). Taken together, these statistics seem to form the perception, at least in the eyes of the jurors and other legal professionals, that a confident witness is more likely to be accurate than a hesitant witness.

The need to emphasize caution in interpreting the relationship between confidence and accuracy is reflected by Wells and Murray's review of over 30 eyewitness studies (1984, cited in Bothwell, Brigham and Deffenbacher, 1987). These authors reported an average correlation between confidence and accuracy measures to be as low as $r = 0.07$. Such low correlations have led some researchers to argue that confidence is not a good predictor of accuracy (Smith, Kassin and Ellsworth, 1989, pp 358). This study by Wells and Murray (1984) and the concerns raised by Smith et al (1989) highlight the importance of investigating witnesses' confidence judgments.

The aim of this chapter is not necessarily to add to the growing debate of the predictive value of confidence on accuracy performance, but to bridge the rather diverse fields of cognitive and forensic psychology. Researchers in the memory and verbal learning fields, use a variety of subjective measures to assess performance. The sub-field of metacognition can be defined in terms of the participant's knowledge about the content, functioning and abilities of their memory system (Costermans, Lories and Ansay, 1992). Some metacognitive processes are measured through the use of feeling of knowing judgments. These are usually applied when the participant fails to retrieve the answer to a general knowledge question, for example. In such instances, participants are asked to rate the probability that they would recognize or 'feel that they know' the correct answer in a subsequent multiple choice test.

A review of the feeling of knowing data by Smith (1989, cited in Smith, Kassin and Ellsworth, 1989) showed that feeling of knowing states are useful predictors of accuracy, producing moderate correlations of $r = 0.55$. Similar correlations have been obtained by Stephenson (1984, cited in Smith, Kassin and Ellsworth, 1989). This particular study required participants to study a short text and then recall various details about the text. For each detail that they recalled they were asked to provide a confidence rating. This study by Stephenson yielded a moderate C-A correlation of $r = 0.48$.

Despite studies from the cognitive domain obtaining moderate and significant correlations for both confidence and feeling of knowing judgments, this pattern is not replicated in eyewitness memory (Smith, Kassin and Ellsworth, 1989). A number of different explanations have been put forward to account for this discrepancy between the cognitive and forensic literature.

One explanation is concerned with the initial research question that is set by the researcher. Researchers from the forensic area tend to focus, almost exclusively on the extent to which confident witnesses differ in their identification performance from hesitant witnesses. Such a question is addressed typically by getting a large number of participants to witness a single (staged or real) crime scene. The performance of participants who give high confidence ratings is compared with participants who give low confidence ratings. This type of comparison is referred to in the C-A literature as between subjects accuracy, because the C-A relationship is determined for the group as a whole.

The cognitive school of thought, on the other hand, is concerned with the variation in accuracy across confidence levels for statements made by the same participant. This type of question is usually addressed by gathering confidence ratings from a participant and calculating the C-A relationship for that participant over a large number of test items. This type of comparison is known as within subjects accuracy, because it is assesses the C-A relationship within a given individual.

One study that adopted the between vs. within subjects approach to the C-A relationship in assessing performance in eyewitness memory and general knowledge tasks was conducted by Perfect, Watson and Wagstaff (1993). These authors showed that for within subjects accuracy, there was not a difference between the two memory tasks. However, for between subjects accuracy a difference between the two tasks emerged, with a strong positive correlation for general knowledge ($r = 0.58$) but not for eyewitness memory ($r = -0.11$). These findings suggest that under some circumstances (within subjects) a participant's confidence level can reflect relative states of knowing. What seems to be lacking in eyewitness memory that is otherwise present in general knowledge memory is the ability to make absolute confidence judgments. Perfect and Hollins (1996) argue that this within vs. between subjects discrepancy between eyewitness memory and general knowledge memory may have something to do with the fact that general knowledge provides participants with an insight into their own strengths and weaknesses. For example, a participant who is an avid sport follower will be able to allocate the relevant confidence category to a question that is sport related than to a question which does not fall into their field of interest, such as geography. This follows that if a geography related question is asked, the participant will be able to correctly use the lower end of the confidence scale to make their judgment.

This anchoring or calibration mechanism that Perfect et al (1993) propose is not available in eyewitness memory as it is difficult to generate a mental checklist relating to person recognition skills. In other words, it is very difficult to verify in eyewitness memory conditions how good we are at remembering someone's facial characteristics and hair colour, for example. As a direct consequence of this, it becomes increasingly difficult to modify the confidence scale in an appropriate and consistent manner — which in turn will result in a low C-A relationship for eyewitness memory.

Other explanations that have been proposed alongside the calibration theory, focus on the qualitative differences between eyewitness memory and general knowledge memory. As Perfect, Watson and Wagstaff (1993) have noted, eyewitness memory is usually based on an event that has only been witnessed once, under non-optimal conditions the events are generally more stressful than tests relating to general knowledge memory. Smith, Kassin and Ellsworth (1989) have added that the events which eyewitnesses are exposed to also tend to be perceptually richer and more complex than in general knowledge memory. Smith et al (1989) provide a further distinction between the two types of memory, in that eyewitness memory may, to some degree contain a confabulatory bias, which is part and parcel of the situational pressures associated with being an eyewitness. This confabulatory bias, can produce the perception that I was there, so I should know , which in turn may encourage the witness to provide details that they can not remember clearly. This confabulatory processing heuristic would consequently impact on the way participants assign the confidence scale to their memories and in turn may lead to the attenuation of the C-A relationship.

Despite the extensive use of confidence judgments by law enforcement agencies and the formal acceptance of these judgments in courtroom proceedings in the U.S Supreme Court (Neil v. Biggers, 1972 cited in Olsson, 2000), the predictive value of confidence on accuracy is worryingly low. The following sections outlines a number of studies that have investigated the C-A relationship in face and voice recognition. The two studies that will form the central focus of experiment 2 in this chapter will include the research done by Olsson, Juslin and Winman (1998) and Yarmey, Yarmey, Yarmey and Parliament (2001).

In the study by Olsson et al (1998) participants were required to listen to a recording of a conversation involving 4 characters engaged in a criminal activity. Participants were later asked to identify each suspect in 4 separate line-ups, one for each suspect. The participant was required to select one voice from a selection of eight voices ( 1 target plus 7 distractors).

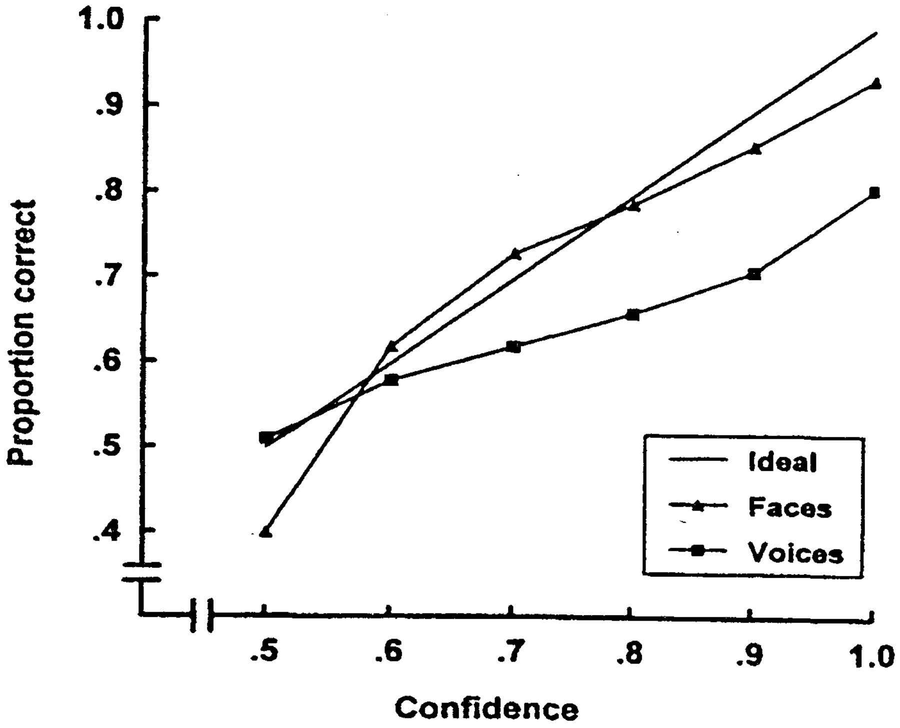
Confidence was assigned through an 11 point subjective (confidence) scale which ranged from 0% (zero probability that the identified person is the culprit) to

100% (1.0 probability, or certainty, that the identified person is the culprit). Olsson et al (1998) argued that participants are considered to be well calibrated in terms of their confidence and memory performance if the subjective probabilities correspond to relative frequencies. For example, across all identifications made in the 100% category, 100% should be correct and so on for the other categories. In this particular study, the researchers explained the concept of subjective to objective calibration to their participants. The aim of Olsson et al's experiment 1 was to collect data from the earwitness domain (voice recognition) so that performance can be compared with an eyewitness study (face recognition) that they conducted previously (Juslin, Olsson and Winman, 1996) .

The overall pattern of results showed that there was very poor correspondence between confidence and accuracy in participant's voice recognition abilities. The pattern of performance observed with voices in Olsson et al's study (1998) seemed to show an over-confidence bias. This was most clearly evident in the 100% confidence category. For this confidence category, participants largely overestimated the probability of a correct identification by producing accuracy rates of around 30% - 40% .

In experiment 2, Olsson et al (1998) tested both voice and face recognition performance in the same sample of participants. In the voice condition, participants listened to a recording of a number of different speakers reading a newspaper article. In the face condition, the speakers' face was photographed and used for study and test. Participants were tested with 64 items, which consisted of 16 original faces, 16 original voices and 32 distractors. In this experiment, participants had to rate their confidence on a six point confidence scale which ranged from 50% (guessing) to 100% (certainty that the correct decision was made).

Figure 4.2a: Overconfidence bias in voice recognition. Figure 3 taken from Olsson, Juslin and Winman (1998).



The findings from experiment 2 showed better performance in the face condition which obtained a proportion of 0.83 correct compared with 0.63 correct for voices. Figure 3 in Olsson et al's paper shows the confidence probability along the x-axis plotted against the proportion correct along the y-axis. Looking at figure 4.2a, it is evident that there is an overconfidence bias in both modalities. However, the degree of overconfidence is greater for voices than faces. In the 100% confidence category, the face condition generates a proportion correct score of approximately 90%, whereas the corresponding value in the voice condition is in the region of 70%. This overconfidence bias for voices persisted in experiment 3.

Olsson et al (1998) initially proposed that the over-confidence bias observed with voices may have been due to the large selection of voices used during the study phase of the experiment, specifically participants were required to encode 16 target voices. Olsson et al (1998) argued that the size of this large selection of voices may have produced interference effects during encoding. This interference effect was observed in Legge, Grosmann and Pieper's (1984) study which demonstrated a rapid decline in performance as the number of studied voices increased. Therefore, Olsson et al (1998) carried out another experiment which required participants to study one face and one voice and were subsequently given two items at recognition. These items consisted of the original stimulus (face / voice) presented along with a distractor stimulus. Even under these conditions, participants continued to display an over-confidence bias for voices at 100% confidence. Performance in the face condition produced perfect confidence calibration, with 100% confidence yielding 100% accuracy. In the 100% confidence category for voices, performance just failed to reach 90% accuracy.

Across all three experiments in Olsson et al's (1998) study, it was shown that overall voice recognition performance is poorly calibrated compared to face recognition. Earwitnesses tend to be poorly calibrated in terms of their confidence and identification performance, even when task demands are reduced. The most notable finding from this study, is that one of the underlying differences between how confidence judgments are assigned to faces and voices may have something to do with voices generating an overestimation bias in confidence. Although both modalities elicit an overconfidence bias, this is most apparent for voices where the discrepancy between 100% confidence and accuracy is in the region of 20% compared with a value a little under 10% for faces (figure 4.2a).

The learning of unfamiliar voices is a difficult skill to master and as pointed out by Van Lancker et al (1987) should not be used interchangeably with voice recognition for familiar voices. Nevertheless, even when encoding conditions are optimal and when the set of voices to be learned is reduced, performance on subsequent recognition tests can still remain problematic (Legge, Grosman and Pieper, 1984). Bearing this in mind, it would be interesting to see what relationship do confidence judgments have in the recognition of familiar voices.

This issue was directly pursued by Yarmey et al (2001) who measured the C-A relationship in voices that varied in different levels of interpersonal familiarity to the listener. In this study, participants were presented with an audio recording of 4 voices which differed in interpersonal familiarity to the participant. They were presented with a voice high in familiarity, such as an immediate family member or best friend, moderate familiarity such as a work colleague and low familiarity such as a casual acquaintance. They were also presented with an unfamiliar voice. On presentation of each voice, participants had to name the voice or say that the voice was unfamiliar to them. For each decision they made, they had to rate their confidence on a 5-point scale, which ranged from 1 (unsure) to 5 (absolutely certain). Each auditory recording gradually increased in duration until the participant made a response. The shortest segment included the single greeting 'hello' and continued gradually upto 2 minutes of continuous speech.

The results from the name data in this experiment produced 85% accuracy for highly familiar voices, which was higher than the percentage of names recalled to moderate familiar voices (79%), which in turn was significantly higher than the percentage of names recalled to voices of low familiarity (49%). There were no significant differences between high and moderate familiar voices.

When the C-A relationship was assessed within each of the familiarity groups, a significant C-A relationship emerged for high and low familiarity with correlations of $r = 0.36$ and $r = 0.56$, respectively. There were no significant C-A correlations observed for voices in the moderate and unfamiliar groups. Therefore for some levels of familiarity confidence is a good predictor of accuracy, whereas for other levels of familiarity it is not.

However, Yarmey et al's (2001) study did not include a face condition, so whilst their data show that their participants can correctly assign confidence judgments to the names of people in response to hearing their voice, is this ability at a comparable level to faces? The experiment designed in this chapter attempts to directly address this issue by looking at occupational and name recall across both modalities.

### 4.3: Confidence and qualitative processes in person recognition memory.

The additional use of recording confidence judgments in this experiment will also serve to assess whether faces and voices are governed by different recognition processes. The idea that recognition memory may be governed by two qualitatively different processes is not new. Mandler (1980) was one of the first proponents to articulate the idea that there may be two processes involved in recognition memory. Mandler's theory was of seminal importance in the psychological literature as it subsequently led to the development of the pioneering Remember / Know (R / K) paradigm by Tulving (1985) and Gardiner (1988). The elegance of the R/K paradigm provided researchers with an alternative way to study these distinct memory processes as originally put forward by Mandler (1980). Furthermore, the qualitative difference between R/K judgments can not be reduced to confidence levels (Gardiner, 2001).

Mandler's (1980) dual-component account of recognition memory incorporated two distinct memory processes. One form of processing was based on inter-event elaborative processes or what is often referred to as contextual retrieval. The other, simpler process, was based on familiarity. To illustrate how these two modes of recognition may operate, Mandler provided the 'butcher in the bus' example. This example was used to show how recognition of a familiar person may come about by recognizing the face as one that has been encountered before (familiarity mode) or by retrieving the relevant context (elaborative processing mode). In this particular instance, the relevant context would be the supermarket, which would be used to produce enough cue specificity to retrieve the memory, in this case, the butcher from the supermarket (Mandler, 1980, pp 252-253). Throughout the recognition memory literature the familiarity mode process continues to be referred to as *familiarity* whereas the elaborative processing mode is referred to as *recollection* — to capture the idea that participants can *recall* certain aspects of the original event (Yonelinas, 2001).

## 4.4: Evidence for a dual process account obtained through confidence judgments.

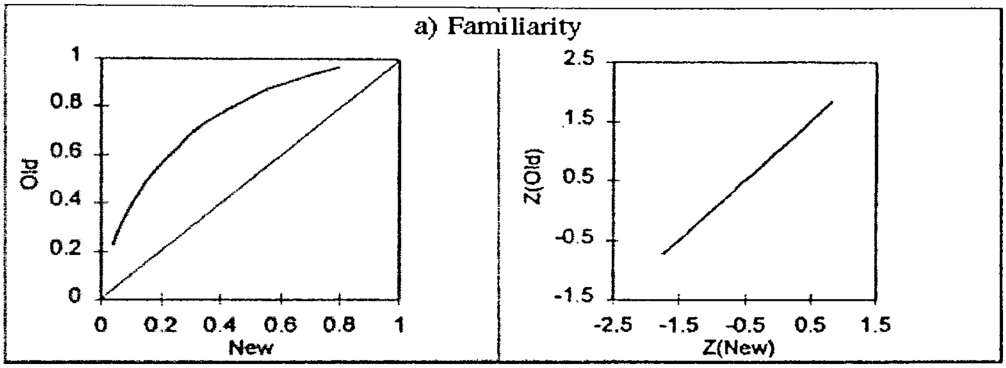
One way of validating the dual-component theory is through the analysis of Receiver Operating Characteristics (ROCs). This technique is particularly useful as it allows researchers to directly assess the relative contributions of these recognition memory processes. ROCs show the relationship between hits and false alarms at various levels of response bias.

Response bias is varied by asking participants to rate their recognition response on a designated confidence scale. Participants are firstly asked if the test item presented to them is 'old' or 'new'. A test item is correctly deemed as 'old' if the item had been presented to them during the study phase of the experiment. If the item is a distractor and therefore did not appear in the study phase, the correct response in the recognition test is to designate the item as 'new'. Once this 'old / new' recognition decision has been made, participants are asked to rate their confidence in this decision.

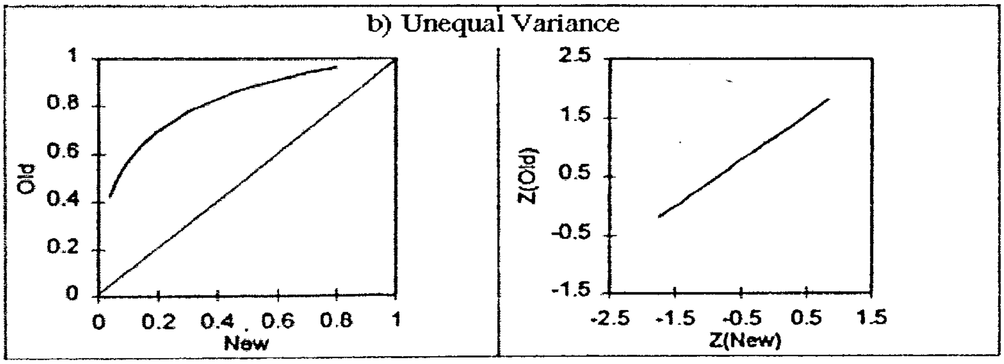
ROC curves are calculated by plotting the cumulative hit rate (old item called old) against the cumulative false alarm rate (old item called new) across the different confidence levels. Conventionally, a six-point confidence scale is used where 6 indicates 'sure the item was studied' and 1 indicates 'sure item is new' (Yonelinas, Kroll, Dobbins and Soltani, 1999). Each point along the ROC curve represents the cumulative probability of hits and false alarms, so that the left most point represents 'sure old' responses to old items (y axis) and to new items (x axis). The second point represents both 'sure old' and 'probably old' responses and so on. This procedure is repeated for all the confidence categories except the least confident category, in which the cumulative hit and false alarm rate would lead to 1.0. Therefore, a confidence scale with N ratings will produce a ROC curve with N-1 points. Once all the points have been plotted, the complete curve can be used to assess the extent to which recall processes may be involved in recognition memory (see figure 4.4a).

Figure 4.4a: Hypothetical ROCs (left panel) and z-ROCs (right panel) for (a) the equal variance familiarity model, (b) the unequal variance model, (c) the recollection model and (d) the dual-process model. Figures taken from Yonelinas (1999).

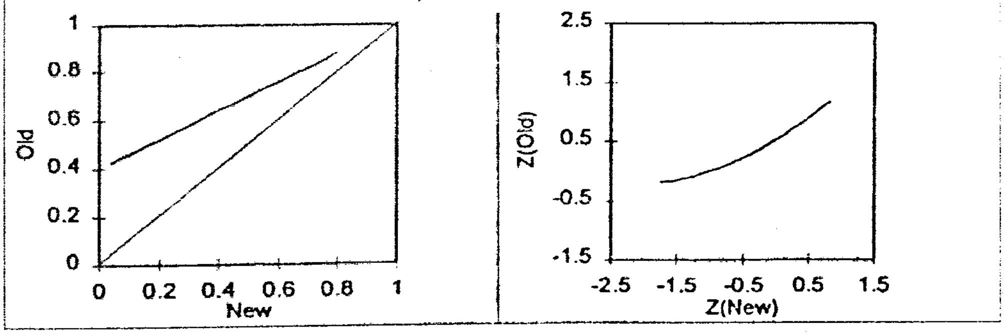
a) Familiarity



b) Unequal Variance



c) Recollection



d) Familiarity and Recollection

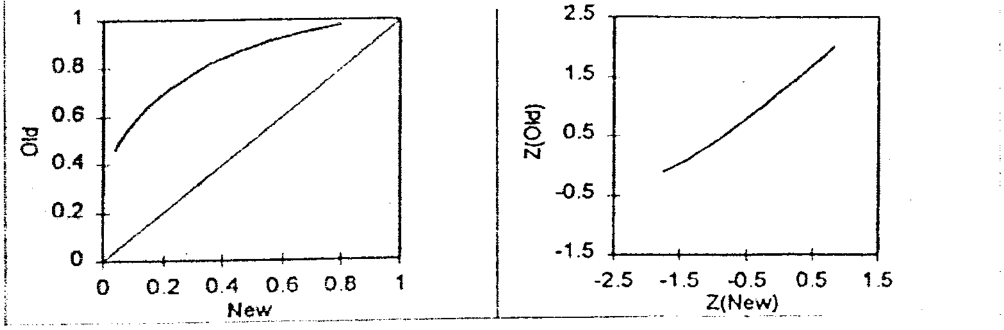


Figure 4.4a is taken from Yonelinas' (1999) paper and shows four hypothetical ROCs (left side figures) and their corresponding curves when plotted in z-space (right side figures). The diagonal line illustrates chance performance. The top figure (a) is an equal-variance familiarity model, which is generally accounted for by signal detection theory (see Yonelinas, 1999). The idea is that participants can place the test items on a familiarity continuum, so that items that were presented during the study phase (old items) fall on the high end of the continuum and new items fall on the low end of the continuum. Since both old and new item distributions overlap, participants must set up a response criterion, so that items only exceeding this criterion are labeled as 'old'. As the criterion is decreased, both hit and false alarm rates will increase smoothly. The ROC generated by this model is characteristically symmetrical along the diagonal line. This is because the old and new items distributions have equal variance. When this assumption is violated, the symmetry of the ROC is destroyed, as shown in figure b, which is referred to as the unequal variance familiarity process model (left panel).

Figure c, is the single-high-threshold recall process, which is generated purely through the retrieval of contextual information and is often termed as recollection. In this case, an item is deemed as 'old' if the participant can retrieve some qualitative aspect of the studied item, if not the item is deemed as 'new', regardless of how familiar it might seem. The model assumes a high threshold process so that 'old' responses to lures and to studied items that are not associated with contextual retrieval are assumed to occur through guessing, which in turn provides the model with its linear component. The proportion of items that exceed this threshold, determines the y-intercept of the ROC curve (Rotello et al, 2000).

The models outlined so far have assumed single recognition models. Figures a and b, represent recognition based purely on assessments of familiarity. Figure c, on the other hand, represents a recognition process that is purely driven by qualitative retrieval of the studied item. Figure d, represents the dual-component model and shows the type of ROC that would be produced if both familiarity and recollection contribute to recognition performance. Figure d, shows a ROC that is curvilinear and asymmetric along the diagonal (left panel). The ROC generates a y-intercept that is greater than zero as produced by the recollection process in figure c. However, it deviates slightly from figure c, by developing a slight curvature, which indicates that familiarity is also contributing to recognition performance. The ROC produced by the dual-component model graphically resembles the unequal variance familiarity process model (figure b) which in turn can make discriminating between these two models very difficult indeed (Yonelinas, 1999).

One way of discriminating between the different models is to replot the cumulative probabilities ROCs using normalized z-scores. The right side of figure 4.4a, shows the z-ROCs for each recognition model. Yonelinas argues (1999) that z-ROCs can be used to differentiate between the various models. If the equal variance assumptions are maintained in the single familiarity process model, then the z-ROC should be linear and have a slope of 1.0. The recollection (figure c) and dual-process (figure d) models produce z-ROCs which are curvilinear. Since the recollection process is assumed to lead to high-confidence hits (Yonelinas, 2001), the curvature should be most pronounced on the left of the z-ROCs. The unequal variance familiarity model (figure b) predicts a curved ROC along with a linear z-ROC (see figure b). Although z-ROCs are a useful means to differentiate between the various models, recent research has failed to replicate the curvature of z-ROCs even under conditions likely to encourage recollection (see Heathcote, 2003), thus challenging the predictions of the dual-process model. For present purposes, both cumulative and z-ROCs will be plotted for each condition and examined to see whether performance in the two modalities can be characterized by one of the four recognition processes outlined in figure 4.4a.

## 4.5: Rationale.

One prediction that is addressed in this study is whether the judgments of familiarity made in experiments 1a through to 1e are interpreted in terms of confidence. If this is the case, then fewer occupations should be retrieved to voices that are given the highest confidence rating relative to faces of the same category. This in turn would replicate the overconfidence bias shown in figure 4.2a in Olsson et al's study (1998) with newly learnt voices. If confidence judgments do not produce any modality effects for occupational performance as observed in the previous chapter, then it may be possible that confidence is based on something quite separate from familiarity. The remainder of this thesis will investigate whether the modality difference can be accounted for by different memory processes.

In order to generate ROCs for faces and voices, the methodology employed in this experiment will deviate slightly from the methodology used in Chapter 3. The same set of stimuli and the same between-subjects design as used in experiment 1d shall also be used in the experiment reported in this chapter. As with previous ROC research, participants will be asked firstly whether the face that they see or the voice that they hear is a famous or unfamiliar person by responding ‘old’ or ‘new’. An ‘old’ decision indicates that the item is famous, whereas a ‘new’ decision will indicate that the item is unfamiliar. For every old/new decision participants make, they have to rate their confidence on a six-point confidence scale as used by Yonelinas et al (1999). These points will be used to plot cumulative ROC plots and z-ROCs for each modality.

Where this methodology parallels the approach used in Chapter 3 and by Hanley et al (1998), is that for every 'old' decision the participant makes, they have to provide occupational details and name the person. By replacing familiarity judgments with confidence judgments it would be possible to investigate whether confidence behaves like familiarity in terms of occupational and name recall. If it does, then there should also be modality difference obtained for these performance measures even when participants are required to make confidence judgments. Furthermore, this should be the most pronounced for voices even at the highest confidence rating, which for this experiment will be 'sure'.

Perhaps the most important aspect of this experiment, is to see whether the use of ROCs and z-ROCs can prove to be a productive research tool in detecting whether there are qualitative and as well as quantitative differences between faces and voices. It may enable us to determine if voice recognition is based more on familiarity whereas face recognition is based more on recollection.

## Method.

### *Experiment 2.*

### Participants.

Fifty participants, the majority of whom were undergraduate students at the University of Essex took part in this experiment. The age range of the participants was 18 to 41 years, with a mean age of 21.6 years (17 male, 35 female). All participants were British and had lived in the UK for a minimum of 15 years prior to testing. All participants had normal to corrected-to-normal vision and normal hearing. All participants were paid £5 for their participation or they were given 1 course credit. Two participants were discarded from the analyses because they failed to follow the instructions. Their data were replaced by two additional participants (1 male and 1 female).

### Apparatus.

The apparatus used in the experiment was essentially the same as that used in the experiments reported in Chapter 3. An Apple Macintosh computer was used to present the stimuli to the participants. Presentation was via the QuickTime Player software program on the Apple Macintosh for the faces condition. The Apple Macintosh's built-in CD player was used for presentation of the voice stimuli.

### Stimuli.

The stimuli used in this experiment consisted of the same items as used in experiment 1d (see Appendix 3). The items were presented in a fixed random order. In total there were 48 famous and 48 non-famous extracts used. Each extract was approximately 7 seconds in duration. Two experimental blocks were created with 24 famous items and

24 non-famous items in each block. This was done to make data collection easier and to allow a short break for the participants in between blocks. Two famous and two non-famous items were used for the practice trial.

### Design.

As with experiment 1d, this experiment used a between subjects design. Twenty-five participants were randomly allocated to the face condition and twenty-five participants were randomly allocated to the voice condition. There were a number of dependent measures taken, including: the number of hits and false alarms in each condition, correct recall of the person's occupation and correct recall of the person's name. All of these dependent measures were looked at in relation to each confidence level.

### Procedure.

Testing took place on an individual basis. Presentation of stimuli was in two blocks of 48 items (24 famous and 24 non-famous). Half the participants completed Block 1 first, followed by Block 2, whilst half the participants did Block 2 first, followed by Block 1. The items were presented in a fixed random order.

In the face condition, participants were seated approximately 30 cm in front of the computer monitor. They were presented with each face extract via the QuickTime Player program. Participants in this condition, only saw the faces without hearing the voices. Presentation of each extract was controlled by the experimenter, by pressing the space bar on the computer keyboard.

In the voice condition, the same seating arrangements were used, however presentation was purely auditory. Each item was played via the CD player in the disc drive of the computer. Participants listened to each extract through a pair of Sennheiser headphones connected to the computer. Presentation of each trial was also controlled by pressing the space bar.

Participants were informed that they would be shown a set of video/audio clips. They were warned that each extract would be brief and presented only once, so they were instructed to pay close attention. The participants were instructed to decide for each extract, whether the item was famous or non-famous. If they thought that the item was famous, they had to indicate OLD and if they thought the item was non-famous they had to respond by indicating NEW on their response sheet. They also had to provide a confidence rating for each OLD/NEW decision on the basis of Yonelinas et al's (1999) 6-point confidence scale (6=sure OLD, 5=probably OLD, 4=maybe OLD, 3=maybe NEW, 2=probably NEW, 1=sure NEW ).

If they made an OLD judgment, participants were required to give a description of the person's occupation, along with any other biographical details (eg, if they are associated with anybody famous, etc). They were instructed not to guess the person's occupation. Finally, they were also required to name the person.

After the first block was completed, participants were allowed a short break before starting the second block of extracts. Once the experiment was completed, participants were de-briefed and paid for their participation. As part of the de-briefing session, each participant was given a list of names of all the celebrities that they heard or saw. They were asked to place a tick next to the name if they were familiar with that celebrity. This task was done to ensure that all the celebrities were well known to the participants.

## Results.

The results section in this chapter is divided into two parts. The first set of analyses will examine the quantitative differences between faces and voices. This part of the results section will be organized in a similar layout as the results presented in Chapter 3. The results section will conclude with the second set of analyses. These will focus on the qualitative differences between faces and voices by comparing Receiver Operating Characteristics (ROCs) between the two modalities.

### *Overall recognition performance.*

Table 4.1.a. Means and standard deviations (in parentheses) for the face and voice conditions.

| | Faces | Voices |
|---|---|---|
| Hits | 44.1<br>(2.7) | 34.4<br>(6.4) |
| False alarms | 6.6<br>(6.1) | 13.5<br>(6.4) |
| Number of occupations correct | 36.8<br>(6.7) | 18.6<br>(6.7) |
| Number of names correct | 31.9<br>(9.4) | 14.6<br>(7.4) |
| Proportion of occupations conditionalized on familiarity | .83<br>(.12) | .54<br>(.16) |
| Proportion of names conditionalized on familiarity | .72<br>(.19) | .42<br>(.19) |

A hit response occurs when a participant assigns an 'old' judgment to a famous item. A false alarm occurs when a participant assigns an 'old' judgment' to a non-famous item. Both mean hits and false alarm values represent a score out of a possible total of 48. A series of independent samples t-test were carried out on the data, which revealed with one exception that all values were significantly higher in the face condition than in the voice condition ($p < 0.05$). The only exception was with the false alarm data, which showed that there were significantly more false alarms to voices than faces ($p < 0.05$). Overall, there were significantly more celebrities recognized in the face modality than in the voice modality. Furthermore, faces were significantly more likely to be associated with occupations and names than voices ($p < 0.05$).

### *Signal Detection Measures.*

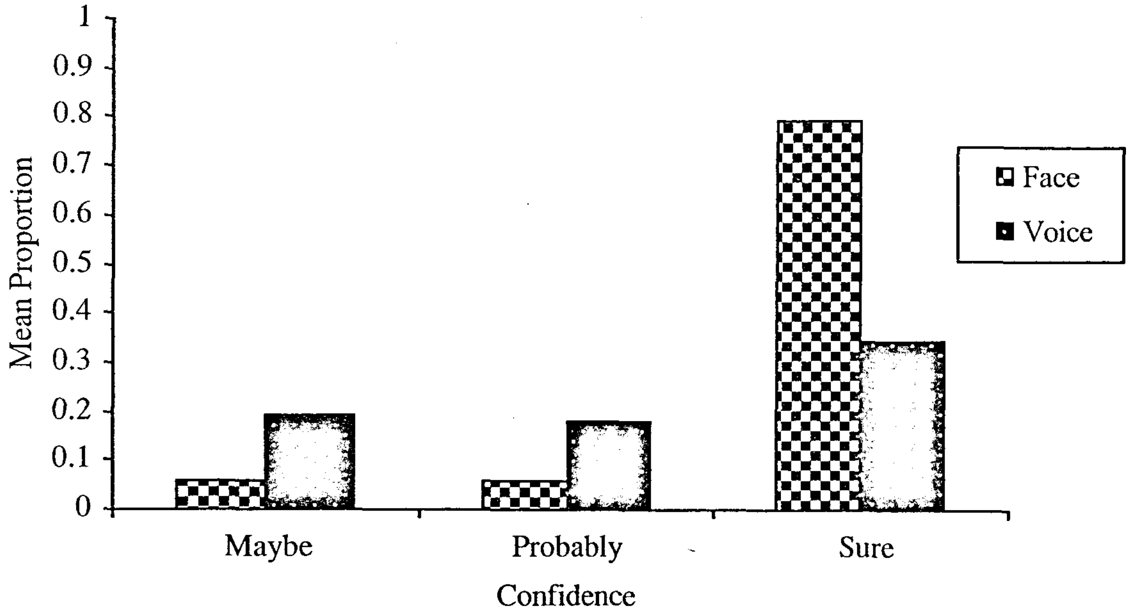
The mean A' value was 0.94 for the face condition and 0.81 for the voice condition, respectively. Both values were statistically above chance ($p < 0.05$). However, the level of discrimination was greater in the face condition than in the voice condition ($p < 0.05$). The degree of bias, as measured by B  D was -0.17 for faces and 0.04 for voices. Neither of these values differed significantly from neutral ($p > 0.05$) or from each other ($p > 0.05$). Even though overall accuracy was significantly better in the face than in the voice condition, the criteria that the participants adopted did not differ between the two modalities.

In order to assess the level of accuracy for the two modalities across the three confidence levels, each participant's score at each level of confidence was converted into a proportion correct score and submitted into a 3 (confidence) by 2 (modality) mixed subjects analysis of variance.

The ANOVA revealed a main effect of modality, F (1, 48) = 51.5, $p < 0.05$ and main effect of confidence, F (2, 96) = 253.7, $p < 0.05$. There was also a significant modality by confidence interaction, F (2, 96) = 100.3, $p < 0.05$. Tukey tests revealed that there were significantly more hits made to faces than voices at the sure category ($p < 0.05$). More hits were made to voices at the lower two categories ($p < 0.05$). The majority of the hits were made at the sure category of confidence which differed significantly from probably and maybe for both modalities ($p < 0.05$). This pattern of performance is displayed in figure 4.1.a. The fact that there was better recognition to faces than voices even at the highest confidence category (sure), illustrates an over-confidence bias in familiar voices as previously observed with newly learnt voices (Olsson et al, 1998).

A similar pattern emerged when a 3 by 2 mixed ANOVA was applied to the overall proportion correct data. This was calculated by the total number of hits given at each confidence level divided by the total recall attempts (hits plus false alarms). Although the main effect of modality was not significant, F (1, 48) = 0.53, $p = 0.470$, there was a significant main effect of confidence, F (2, 96) = 73.0, $p < 0.05$ as well as a significant interaction, F (2, 96) = 9.90, $p < 0.05$. Performance for both modalities improved consistently with each increase in confidence ($p < 0.05$). However, at the highest confidence category (sure), performance was significantly better in responses to faces than voices. A planned comparison t-test showed that the overall accuracy value for faces (0.99) was significantly higher than the accuracy value for voices (0.85), $p < 0.05$.

Figure 4.1.a. Proprtion of Hits (OLD) made at each level of confidence for faces and voices.

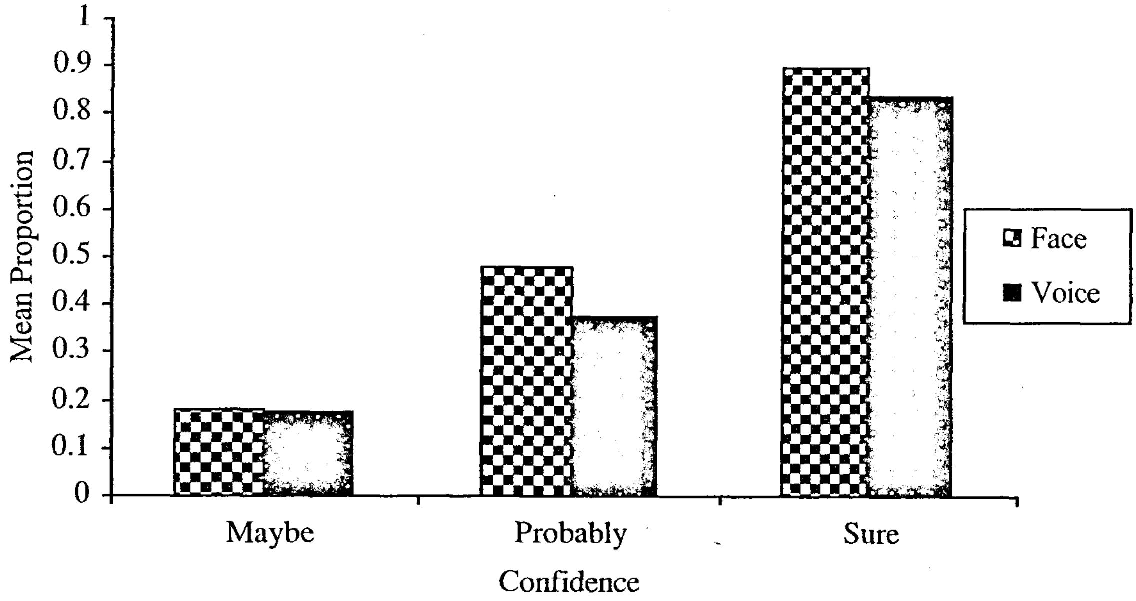


*Retrieval of person-specific identity information.*

*Occupational recall.*

The same marking criterion that was used in Chapter 3 was applied in this experiment. Participants were always probed for detailed occupational descriptions about the person. If they could only retrieve generic information, such as 'TV presenter', their response was classified as a familiar-only. Occupational retrieval was conditionalized on each level of confidence. The data were subjected to a 3 (confidence) by 2 (modality) mixed ANOVA. The analysis revealed that there was no effect of modality F (1, 48) = 1.64, p = 0.205, but a significant effect of confidence F (2, 96) = 161.8, p < 0.05. The interaction was not significant, F = (2, 96) = 0.85, p = 0.43. An analysis by items also failed to produce a main effect of modality, F (1, 94) = 1.09, p = 0.300. This pattern of results is displayed in figure 4.1.b.

Figure 4.1.b Proportion of Occupations recalled at each level of confidence for faces and voices.



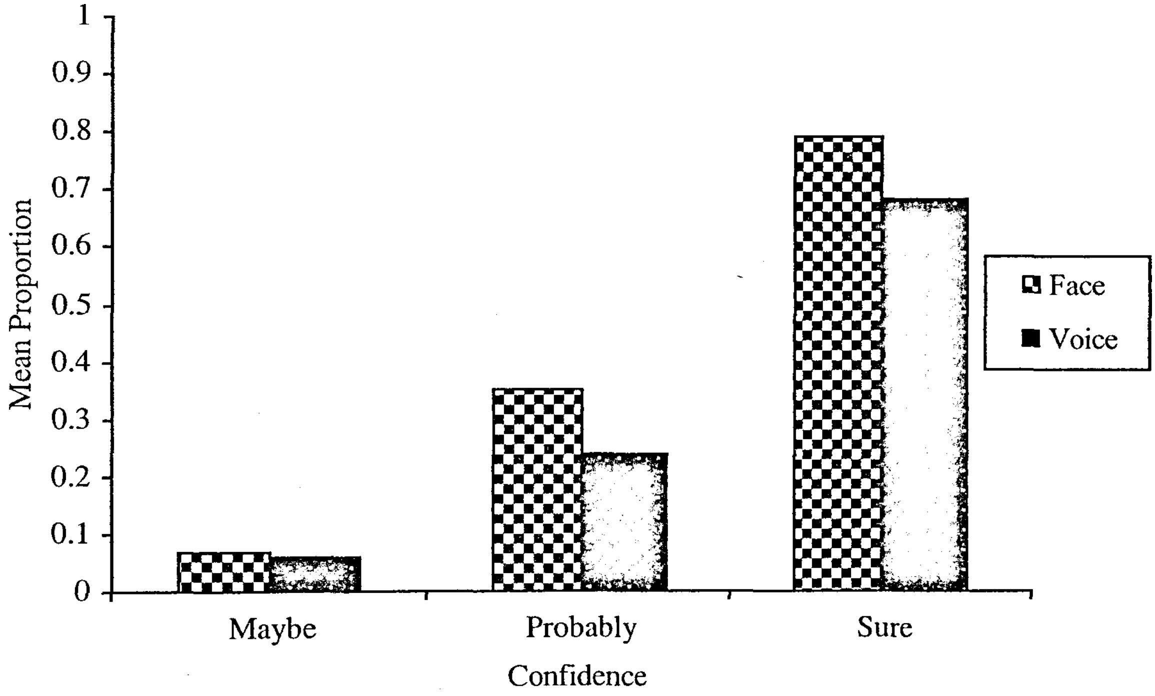
Although there were fewer voices than faces recognized overall (table 4.1.a), for each level of confidence the proportion of occupations recalled was the same for both modalities. It appears that familiarity judgments can be dissociated from confidence judgments. In the case of the latter, a very confident response to a face is just as likely to be interpreted as a very confident response to a voice.

*Name recall.*

A 3x2 mixed ANOVA was also applied on the proportion of names correctly recalled as conditionalized on each level of confidence. The analysis with the name data extends the findings from the occupational data. Here again there was no main effect of modality F (1, 48) = 2.90, p =0.09, but there was a main effect of confidence F (2, 96) = 178.4, $p < 0.05$. The interaction was not significant, F = (2, 96) = 1.34, p = 0.27.

This suggests that even with name recall, participants are just as efficient at retrieving the person's name to the celebrity's face as they are to their voice (see figure 4.1.c) when confidence levels are matched.

Figure 4.1.c. Proportion of Names recalled at each level of confidence for faces and voices.



*Receiver Operating Characteristics.*

Receiver Operating Characteristics (ROCs) were created by plotting the cumulative probability that a famous item will be categorized as 'old' (the hit rate) on the y-axis against the cumulative probability that a non-famous item will be categorized as 'old' (the false alarm rate) on the x-axis across the different confidence levels. For each modality, ROCs were plotted both in probability space and z-space (see figure 4.1d and 4.1e) and the shape of each curve was examined to assess the relative contributions of recollection and familiarity (Yonelinas, 2001).

*Faces.*

The fact that performance was approaching ceiling levels in the face condition (see table 4.1.a) is clearly illustrated in Figure 4.1d. The top figure shows the standard ROC for the face condition, which is clearly well above the diagonal line (chance performance). Looking at this ROC alone, it would appear that it closely resembles the ROC generated from an unequal variance familiarity model and the dual-component model (see figure 4.4a). However as pointed out by Yonelinas (1999), an analyses of the ROCs in z-space is also needed in order gain a clearer picture of the different processes that may be contributing to performance. If face recognition is driven by an unequal variance familiarity process, then according to Yonelinas (1999) the ROCs produced in z-space should be linear.

A close inspection of the bottom figure in figure 4.1d does not appear to be consistently linear. Furthermore, there seems to be a slight upward curve towards the leftmost part of the z-ROC. In order to assess the linearity of the z-ROC, a linear regression analysis was applied on the z-ROC. The $R^2$ from this analysis was then compared with a regression analysis that contained a quadratic component. If the $R^2$ value is greater for the quadratic component that the linear one, then it would suggest that the z-ROC for can be described as curved rather than linear.

The linear regression produced the following equation for the z-ROC, z(hits) = 0.3862 Z(FA) + 1.8199. The proportion of variance accounted for by this model was $R^2 = 0.9887$. The quadratic component produced the equation, z(hits) = 0.038 Z $(FA)^2$ + 0.4854(FA) + 1.8513. Including the quadratic component raised the $R^2$ value to

0.9959. However, the variance account for by this additional component was not significant, $F_{INC}(1, 2) = 3.46$, $p > 0.05$. Therefore, the process most likely to be associated with face recognition is the unequal variance familiarity model.

*Voices.*

The ROCs and z-ROCs for the voice condition are shown in the top and bottom half of figure 4.1e, respectively. The ROCs for the voice condition shows considerably more curvature than in the face condition. However, it is not clear that the level of curvature is entirely symmetrical along the diagonal. A ROC that is completely symmetrical along the diagonal would indicate a recognition process that is based on the equal variance familiarity model (signal detection model). If this is the case, then by transferring the ROC into z-coordinates should, according to Yonelinas (1999), produce a slope of 1.0. The linear regression that was applied to the z-ROC produced the following equation, z(hits) = 0.8206 Z(FA) + 0.9625, $R^2 = 0.9917$. The fact that the slope 0.8206 was not 1.0 (CI: 0.682 – 0.959), suggests that the equal variance model is not an adequate model to capture the data.

A close inspection of the z-ROC also shows a slight element of curvature, albeit towards the far right area of the curve. The fact that this curvature was displaced in a different location than was initially anticipated, a quadratic regression was not applied to the data. Furthermore, the $R^2$ value of 0.9917, indicates that a linear fit can adequately describe the data. With a slope value less than 1.0, the voice z-ROCs are not consistent with the equal variance familiarity model. The only alternative model that can capture the data, is the unequal variance familiarity model (see figure 4.4a). This type of recognition model predicts ROCs which are curved but not symmetrical along the diagonal – a characteristic exhibited by the ROCs for voices. Furthermore, the z-ROCs for an unequal variance familiarity model predicts linear z-ROCs with slopes *not* equal to 1.0 - this assumption is also met by the z-ROCs for voices. Therefore, it appears that the recognition process that contributes to voice recognition is adequately captured by the unequal variance familiarity model.

Figure 4.1d: Average ROCs (top panel) and average z-ROCs (bottom panel) for the face condition. Diagonal line represents no discrimination.

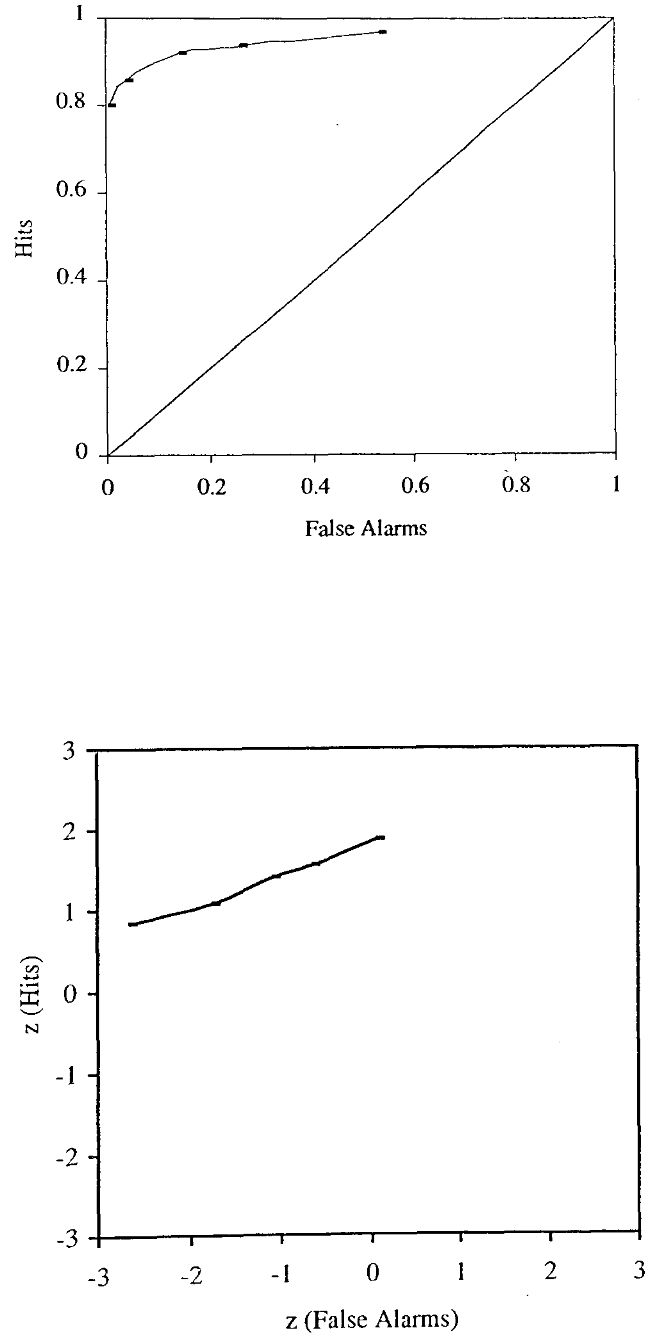
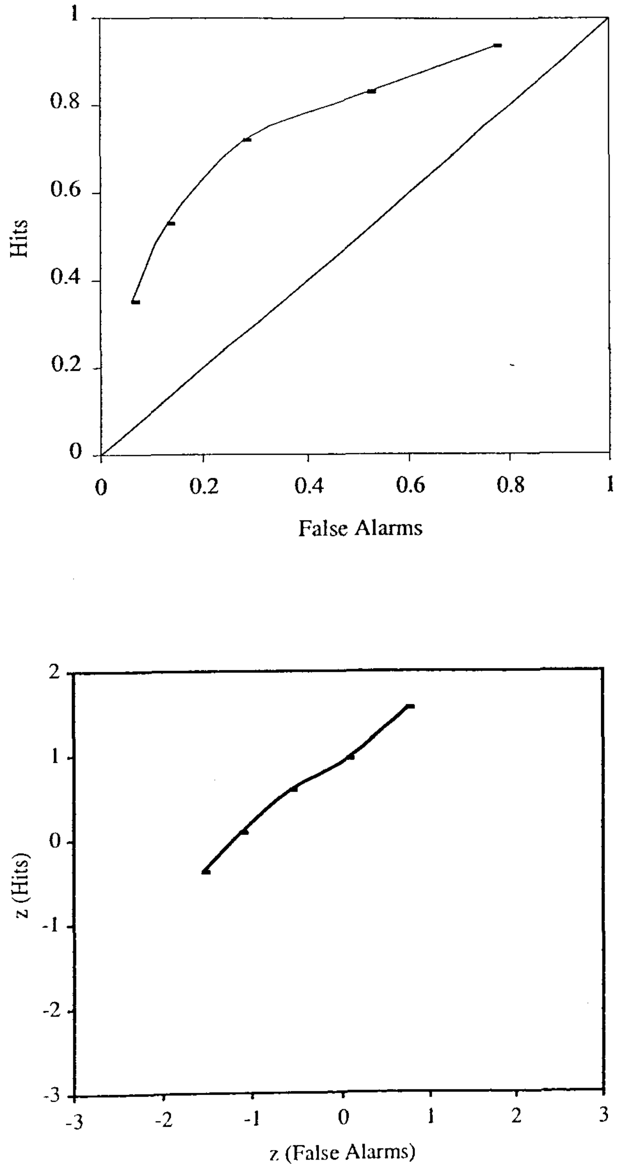


Figure 4.1e: Average ROCs (top panel) and average z-ROCs (bottom panel) for the voice condition. Diagonal line represents no discrimination.



## General Discussion.

Previous research has shown that when a face is rated as being high in familiarity, it is considerably more likely to elicit correct contextual information than a voice of the same familiarity rating (Hanley et al, 1998, Chapter 3). The experiment reported in this chapter, attempted to extend these results by replacing judgments of familiarity with judgments of confidence. Do participants treat these two judgments as meaning the same thing?

Research with confidence judgments conducted on unfamiliar faces and voices has shown that participants tend to show an overconfidence bias to voices relative to faces. In one of Olsson et al's (1998) experiments, the proportion correct for the highest confidence category was approximately 0.90 for faces compared to 0.80 for voices. In a simplified version of their experiment, confidence calibration for the face condition was perfect (100%), whereas accuracy in the voice condition remained at the same 0.80 level. This discrepancy in accuracy for the highest confidence category, was taken by Olsson et al (1998) as a measure of the overconfidence bias, which was more pronounced with voices than with faces.

Confidence calibration was measured in this experiment by analyzing accuracy performance for the two modality conditions across the three confidence levels. The first set of analyses simply included the overall hit rate, which showed that there were significantly more hits made to faces than voices even for 'sure' confidence judgments. This pattern persisted when accuracy was measured in terms of the proportion of recall attempts made. As with Olsson et al's (1998) study, participants were very well calibrated in terms of their confidence and accuracy for faces, reaching almost perfect calibration (0.99) at the sure category. For voices, confidence calibration was quite high (0.85), however it was significantly lower than face recognition performance. Therefore, it appears that even when participants are required to make confidence judgments to familiar voices, an overconfidence bias is present just as it is for newly learnt voices (Olsson et al, 1998).

As well as looking at the overall recognition rate, participants were required to provide the correct occupational description and name to each face / voice that they had previously recognized. In the previous chapter, it was shown that faces are more likely to elicit biographical information than voices, even when overall familiarity is equated. Yarmey et al's (2001) study showed that at some levels of familiarity, confidence provides a good measure of accuracy for name recall. Yarmey et al's study (2001) only assessed name recall to voices that varied in interpersonal familiarity to the participants – a face condition was not used. The experiment here, used a total of 48 famous and 48 non-famous faces and voices and found no evidence of a modality effect for both occupational and name performance (see figures 4.1.b and 4.1.c) when participants were required to base their recognition on confidence judgments. Therefore it appears that judgments of familiarity are not the same as judgments of confidence.

However, this conclusion can only be made tentatively. This is because the only parallel between this experiment and the experiments conducted in Chapter 3, is that participants were always required to provide biographical details to the faces / voices that they previously recognized ('old' responses). The fact that the response scale used in this experiment differed vastly from the one used in Chapter 3, may have altered their pattern of responding.

In this experiment, participants were required to separate their responses to indicate that they recognize the item ('old') or not ('new'). This stage of the response scale is comparable to the familiar / unfamiliar category used in Chapter 3. However, where the two response scales differ is that participants in this experiment were required to judge their 'new' responses on the same scale as their 'old' responses. In the experiments reported in the previous chapter, participants were required to give additional ratings only for items that they recognized (familiar). Therefore, including three additional response categories may have affected the participants' decision criteria.

Previous work that has manipulated testing procedures adds weight to the argument outlined above. For example, Eldridge, Sarfatti and Knowlton (2002) showed that even subtle changes in testing procedures can affect the proportion of hits and false alarms made to remember and know responses. In this particular word recognition study, participants were either instructed under the one-step procedure or the two-step procedure. In the one step procedure, participants were asked to rate each word by giving a Remember (R) or Know (K) response. They were also required to give a New response (N), if they thought that the word was not presented during the study phase. Participants under the two-step procedure were also required to make R, K and N responses – however they were instructed to do this *after* they had decided whether they recognized the word or not.

The two subtle testing procedures affected the number of R and K responses reported by the participants. The number of K false alarms for the one-step condition exceeded the K false alarms for the two-step group. The number of R responses made in the one-step group was significantly greater than in the two-step group. These differences occurred even though overall recognition between the two conditions did not differ.

This research by Eldridge et al (2002) illustrates how different response instructions can strongly influence performance on a subsequent recognition task.

With this in mind, it would be prudent to replicate this experiment by using Hanley et al's (1998) original response scale for assessing confidence judgments across the two modalities.

The main reason for departing from the original response scale used in Chapter 3 was to conduct a separate set of analyses to examine recognition processes across the two modalities. One particular way this can be done is by creating ROCs for both modalities by using the cumulative probabilities of hits vs false alarms across the different confidence intervals.

The top figures of 4.1d and 4.1e show the ROCs for the face and voice condition, respectively. By comparing the ROC for the face condition with the hypothetical ROCs presented in the introduction, the shape of the ROC for faces appears to resemble the ROC generated by the unequal variance familiarity and the dual-component model. According to Yonelinas, (1999), if a ROC is generated by an unequal variance familiarity model, then the corresponding z-ROC should always be linear (see Yonelinas, 1999, page 1419). When the ROC values were normalized into z-scores and replotted (bottom figure), the z-ROC yielded a slight curved component at the left side of the curve. However, statistical analyses showed that this level of curvature was not significant. The findings suggest that face recognition is achieved by assessing the levels of familiarity associated with the test item.

Although a dual-component account of face recognition would have provided an elegant account of the person identification advantage for faces over voices (Chapter 3), there are a few theoretical problems that need to be considered when interpreting the ROCs results.

As noted by Yonelinas (2001) ceiling and floor effects can cause problems for the way in which ROCs are interpreted. If ROCs include points with a hit rate approaching 100% or a false alarms nearing 0%, then this can artificially distort the ROC shape. This also causes problems for the subsequent z-ROC as z transformation is undefined for probabilities of zero or one. A value of 0% for false alarms in the highest confidence category was made by 23 out of the 25 participants tested in the face condition. This would have resulted in 23 missing values for the generation of the z-ROC. As a result, the points for the ROC and z-ROC plots reported in this study were made by averaging the hit and false alarm values across the participants. However, this technique can also be quite problematic (Yonelinas, 2001).

One way in which this floor / ceiling issue in face recognition can be resolved is by reducing the overall recognition of faces. Previous work has shown that by presenting faces in a blurred format can reduce face recognition performance to a comparable level to voices (Hanley and Turner, 2000). Other face degradation techniques that have been studied in the face recognition literature include inverted and scrambled faces (Collishaw and Hole, 2000) and photographic negatives (Lander, Christie and Bruce, 1999). The use of one of these degradation techniques could lead to a decrease in hit rates coupled with an increase in false alarm rates. This in turn should control for any artificial distortion of the ROCs and z-ROCs. It would be interesting to see whether a dual-component model of face recognition would be found under this particular manipulation. This is a particularly relevant research question to pursue given the findings from previous research which suggest that differences in accuracy do not necessarily affect the contributions of recollection and familiarity (Yonelinas, 2001, Yonelinas, Kroll, Dobbins, Lazzara and Knight, 1998).

Figure 4.1e (top panel) shows the ROC generated for the voice condition. Since overall performance was not approaching ceiling levels, the ROC shape was easier to observe. The general shape of the model did not appear to be sufficiently symmetrical along the diagonal in order to associate the model with an equal variance familiarity model.

This means that the initial recognition process in the voice modality, as in the face modality, is based purely on assessments of familiarity, so that the most familiar items are recognized in the absence of any recollective information. However, the asymmetry in the ROC is caused by the old (famous) and new (non-famous) items having unequal variance.

A final point that needs to be addressed in more detail is the shape of the voice z-ROC. As it has already been briefly mentioned, the z-ROC for voices showed curvature towards the right side of the curve. Yonelinas' (1999) ROC account does not propose such deviations from linearity (see Figure 4.4a) However, it is worth to note that such non-conventional curves have been observed by previous researchers. (Ratcliff, McKoon and Tindall, 1994, Glanzer, Hilford, Kim, and Adams 1999a, 1999b, Rotello, Macmillan and Van Tassel, 2000). Rotello et al (2000) suggest that this 'dip' in the ROC and z-ROC is produced when participants use a guessing strategy. This is most likely to occur for the 'maybe old' and 'maybe new' categories, whereby participants find it difficult to assess the level of familiarity for these items. As a consequence, participants are likely to guess equally often for maybe old and maybe new responses. This in turn would lower the middle points in the ROCs. This particular interpretation suggests that the dual-process model as it currently stands needs to be modified to incorporate a third, guessing process. Indeed, as Yonelinas (2001) points out the dual-process model with its two recognition processes may be too simple and acknowledges the need to modify the model to incorporate a guessing or noise process.

Another criticism of the dual-process model comes from the assumptions made by the ROC and z-ROC analysis. The main advantage of creating z-ROCs is to discriminate between the unequal variance familiarity model and the dual process model. As both types of models generate similar ROCs in probability space (see figure 4.4a), z-ROCs are used to differentiate between the two models. This is done by examining whether the z-ROC shows an upward curve towards the left most part of the z-ROC. If this is found, then it can be concluded that recollection is also contributing to recognition performance (Yonelinas, 1999). The shape of the z-ROC is therefore integral to the predictive validity of the dual-component model.

Heathcote (2003) carried out four experiments to test this particular assumption of the dual-component model by generating a series of z-ROCs for various word learning experiments. Across the different word learning experiments, Heathcote (2003) failed to find the characteristic upward curve in his z-ROC analysis. The test items used by Heathcote (2003) included highly similar targets and lures, thus ensuring that recollection was more likely to contribute to recognition performance than familiarity (Rotello et al, 2000). However, the z-ROCs failed to show an upward curve, therefore questioning Yonelinas' (1999) claim that z-ROCs provide useful, converging evidence for the dual-process account.

Another existing problem with the ROC approach to understanding dual-component recognition processes rests on the argument of whether the ROC approach is psychologically valid. The ROC approach can summarize quantitative data, but to what extent can it provide a window on the psychological processes involved?

According to Yonelinas (2001) the ROC approach to recognition memory should be used in conjunction with other research methods such as the Remember / Know procedure (Tulving, 1985). Whilst the ROC approach attempts to provide a quantitative account of recollection and familiarity (Yonelinas, 1999), the R / K approach attempts to differentiate between the two components in terms of variations in conscious experience. Both approaches focus on different aspects of recollection and familiarity, however as argued by Yonelinas (2001), they appear to capture similar underlying processes. The core argument that is relevant to both approaches is that recognition judgments are governed by two distinct retrieval processes rather than a single familiarity signal detection process. However, since both approaches differ in the way in which they account for recollection and familiarity, there are important differences in how these approaches measure the two processes. With some of the inconsistencies observed in the current experiment with the ROC analysis and the recent failure to replicate z-ROC shapes (Heathcote, 2003), the R / K paradigm will be applied to face and voice recognition performance in the following chapters to see whether this approach can provide a better description of the data.

To conclude, the findings from this study represent a unique attempt at assessing confidence judgments in familiar face and voice recognition. Studies that have compared both familiar face and voice recognition are few and far between. The main method that is generally employed by such studies is to assess the level of accuracy at varying levels of confidence. In Olsson et al's study (1998), confidence judgments to voices were not a good predictor of accuracy. Specifically, Olsson et al showed that even at the highest confidence level accuracy for voices was poor relative to faces. This means that voices are more likely to exhibit an overconfidence bias than faces. The findings from the current experiment extends this observation with famous faces and voices by showing that for the highest confidence category (sure), participants' accuracy was 0.85 for voices and 0.99 for faces – a difference which was significant.

The other area of interest in this experiment was to see whether familiarity judgments are assessed in terms of confidence. If so, then participants should also recall fewer occupations and names to voices than faces even when confidence levels are matched. The results failed to produce any modality effects between faces and voices for both occupational and name recall. This may suggest that decisions about familiarity do not operate in the same way as decisions about confidence in relation to biographical recall to famous faces and voices. However, such a conclusion is made cautiously given the discrepancy in the testing procedures used between the experiment reported here and the ones conducted in Chapter 3. This particular point is reinforced by the findings made by Eldridge et al (2002) who found qualitative differences in recognition performance when testing procedures were subtly altered.

A further aim of this experiment was to compare familiarity and confidence judgments across experiments and to generate ROCs from these confidence judgments. Previous research has shown that this technique alongside z-ROCs can be used to assess the relative contributions of recollection and familiarity (Yonelinas, 2001). The ROC analysis for faces proved difficult to interpret given the ceiling performance levels. Although a recollection component may have contributed to face recognition, the levels of performance in the face condition may have masked the extent of this process. What is obviously needed is a comparison of ROCs between faces and voices when overall levels of recognition in the face condition are reduced.

The interpretation of the ROCs and z-ROCs analyses from the current experiment was not conclusive on a number of different grounds (Heathcote, 2003, Rotello et al, 2000). Therefore, the remaining chapters of this thesis will investigate whether differences in recognition processes can be more sufficiently captured by a different research approach. Specifically, Chapter 6, will examine whether the R / K procedure can be applied to the recognition of famous faces and voices.

# Chapter 5: Matching face and voice recognition.

## 5.1: Introduction.

Throughout the experiments reported so far, overall recognition has consistently been better in response to faces than voices. In other words, faces have generated significantly more hits and significantly fewer false alarms than voices. This is likely to reflect the fact that the face of a famous person is more likely to be encountered in the media (newspapers, magazines, television, etc) than their voice. This high level of performance in the face condition made the ROCs and z-ROCs produced in the previous chapter difficult to interpret. Specifically, it was difficult to conclusively argue for a recall component contributing to recognition in the face condition.

## 5.2: Rationale.

The aims of this chapter are two-fold. The first aim is to provide a testing condition for faces that elicits similar overall levels of performance to voices. This will be acquired by applying various degradation techniques to the selection of faces that have so far been used in this thesis and analyzing their performance measures to determine which technique produces recognition performance at a comparable level to voices. This new 'face' condition will be used alongside the selection of voices used throughout this thesis and the non-degraded faces, which from this point on will be referred to as the 'standard' face condition. These three modality conditions will be used in the Remember / Know study in the final experimental chapter of this thesis, with particular emphasis on the distribution of these memory responses when overall accuracy between face and voices is matched.

The second aim of this study is to examine whether the contributions of recollection and familiarity will vary as overall accuracy is reduced. This will be analyzed through the generation of ROCs and z-ROCs. This will therefore extend the work carried out in the previous chapter by analyzing the contributions of recollection and familiarity in face recognition across the different degradation techniques. As a reminder, the ROCs and z-ROCs for the face condition did show some evidence of a recall component, however this was not supported statistically. It was argued that the overall levels of performance in the face condition may not have been sensitive enough to significantly yield a recall component. Therefore, it would be reassuring to see whether recall contributes to face recognition when ceiling levels are removed.

The following sections provides a brief literature review of the various degradation techniques that have been used with moving faces and which will be applied on the standard faces in this pilot study.

### *a.) Inverted Faces.*

Presenting faces in an upside down format has a long-standing history in the face recognition literature (Yin, 1969). This face presentation format often serves as an experimental control. This is because the stimulus complexity, contrast, brightness and configurational properties in an upside-down / inverted face are the same as in a normal orientated face (Valentine, 1988).

The configuration of a face consists of the spatial relationship between the local features of the face such as the eyes, nose and mouth (Collishaw and Hole, 2000). Therefore any differences in speed and/or accuracy between processing these two items can not be attributed to stimulus factors, but instead to factors affecting the observer – so that inverted faces are processed differently from normal faces. One hypothesis that has been put forward to account for the inversion effect is that the configural processing of faces is markedly disrupted through inversion (Sergent, 1984). The fact that this inversion effect is more pronounced for faces than other objects (eg, houses and aeroplanes), is often taken to support the argument that the cognitive processing of faces is somehow a 'special' process (Yin, 1969). Such an assertion has accumulated a wealth of research into the ongoing debate whether faces are 'special' or not (Kanwisher, 2000, Valentine, 1988). The debate is picked up again in a recent review by Liu and Chaudhuri (2003) and will not be reviewed further.

Whilst earlier studies looking at the effects of inversion on face recognition have generally used static images, it is worth considering what role if any does movement play in the recognition of inverted faces? According to Knight and Johnston (1997) movement only has a facilitative effect for negative faces (see section c), but not inverted faces. However, in a follow-up study Lander, Christie and Bruce (1999) managed to show that movement does indeed have facilitative effects for inverted faces. For the moving inverted face condition, the percentage of names correctly recalled was 61% compared with 48% for a multistatic condition whereby three static images of a famous person's face were presented. This led Lander et al (1999) to argue that under non-optimal conditions, movement may enable the viewer access to a stored representation of the person's face.

For the current purposes of the experiment, it was expected that this inversion effect would move overall face recognition performance away from ceiling levels. However, what was central to the current aims of the study was whether this particular degradation technique would result in a level of overall performance that is similar to voice recognition.

### *b.) Blurred Faces.*

In contrast to inverted faces, the removal of the high spatial frequencies in a facial image (blurring) is more likely to disrupt the processing of facial features (eyes, nose, mouth) than its configuration (Collishaw and Hole, 2000). As the level of blur increases the degree of featural detail decreases, so that that the fine-detail outlines of the nose and mouth become progressively coarse, eventually inhibiting facial recognition. Despite the decrease in featural information, according to Collishaw and Hole (2000) recognition can still be achieved by analyzing the configuration of these coarse features. However, at the highest level of blurring, the information that is left in the facial image is too coarse to allow for the analysis of even the most general patterns of the face.

There are different ways in which the blur technique can be applied to faces. One way is to feed the images into a video projector and alter the projecting lens so that the image is out of focus (Hanley and Turner, 2000). Recent developments in multi-media software has allowed researchers more control over this technique by adjusting the placement and level of blur digitally (Collishaw and Hole, 2000, Lander, Bruce and Hill, 2001)

This particular degradation technique has sparked renewed interest on both theoretical and practical grounds. From a theoretical perspective, the advantage this technique has over other degradation techniques such as inverted and negative faces is that it represents perhaps a more naturally occurring form of facial degradation. As pointed out by Lander et al (1999), a blur effect that is applied experimentally is likely to simulate the perception of seeing a face from a distance. From a practical perspective, applying a blur effect on facial images is a common technique used by television programme producers to mask and protect the identity of witnesses shown on television interviews. Even for this technique, movement provides additional benefits to aid recognition. For the most severely degraded images, movement provided a clear recognition advantage over static images. However, for the least degraded images, movement did not show a significant improvement over the static images (Lander, Bruce and Hill, 2001).

Furthermore, blurred faces as used by Hanley and Turner (2000) has shown to be effective in reducing face recognition to a comparable level to voice recognition. With this in mind, it was anticipated that this degradation technique would also elicit similar levels of performance as observed in the voice condition in experiment 2 (Chapter 4).

### *c.) Negative Faces.*

Shading provides important information with regards to the three-dimensional (3D) structure of a face. These depth cues can be disrupted by presenting faces as photographic negatives. This works by reversing the sign of the brightness contrast between image points but retains the 2D shape of the face and the features within the face (Johnston, Hill and Carman, 1992). The lack of depth cues is subsequently restored through movement, as illustrated by Knight and Johnston (1997) who observed improvement in accuracy with moving negative images (61.4%) compared with stills (42%). This finding was also replicated by Lander et al (1999) who expanded the static display to include additional still images. It therefore seems that movement plays two main roles in facilitating face recognition. One obvious contribution that is made by movement is that it provides the viewer with more images of the face. However, this is not the only beneficial effect of movement. As pointed out by Lander et al (1999), it is the way in which a face changes over time –

the *dynamic* information about a face that is integral to the beneficial effects of movement.

Unlike inverted faces that have been shown to disrupt configural processing and blurred faces which hinder the processing of facial features, the type of processing that is involved and / impaired in the recognition of negative faces is unclear. According to Collishaw and Hole (2000) normal face recognition is achieved through the additive effects of configural and featural processing. If one of these routes is disrupted, recognition can still be achieved by accessing the alternative route. If both processing routes are unavailable, then recognition is severely impaired. Therefore inverted faces can still prompt recognition because the facial features are available for analysis. Similarly, blurred faces will allow for the face's configuration to be adequately analyzed despite the lack of featural information. When both inversion and blurred techniques are combined, then access to these routes is markedly impaired thus resulting in poor recognition performance. Indeed, this was what was found by Collishaw and Hole (2000) who found the lowest level of accuracy when these two techniques were combined than when they were applied separately.

Since it may be possible that both configural and featural information are retained in negative faces, it would be interesting to examine the kind of performance measures elicited by this degradation technique and whether these measures are at a comparable level to voices.

## Method.

### *Experiment 3.*

### Participants.

Participants were thirty visit day students and their families who visited the Department of Psychology in July 2002. The age range of the participants was 16 – 41 years with a mean age of 20.6 years. There were 26 female and 4 male participants. All were native speakers of English and had lived in the U.K for the past 15 years. All had normal to corrected vision and normal hearing. The sample of participants were randomly allocated into three equal groups of 10 participants per group.

### Apparatus.

Stimuli presentation was via the QuickTime Player on the Apple Macintosh computer.

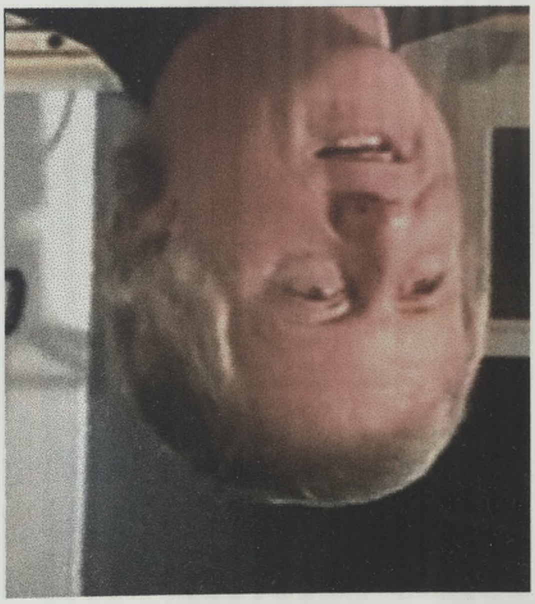
### Materials.

The selection of stimuli used in this pilot study was the same as the selection used in the face condition in experiment 2 and experiment 1d. However, each face was subjected to a degradation technique, which included presenting the faces in an inverted, blurred and photographic negative format. These three techniques served as the three levels of the independent variable. A description of the degradation techniques along with an example is provided in the sections below.

### *Inverted Faces.*

A total of 48 famous and 48 non-famous video clips were used. The stimuli were exactly the same as the ones used in the previous chapter, however each clip was rotated through 180° using the image editing tool on Final Cut Pro on the Apple Mac. An example of an inverted face is displayed in Figure 5.a below.

Figure 5.a: Chris Tarrant, presenter of ‘Who wants to be a Millionaire’. Image rotated through 180° using Final Cut Pro.



### *Blurred Faces.*

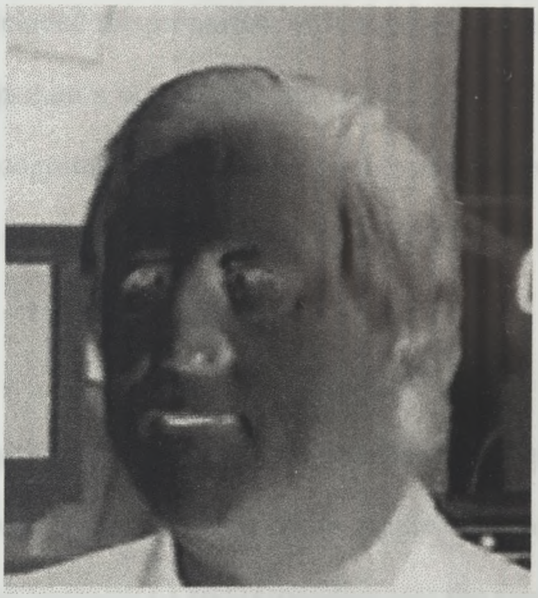
Each face was modified through a Gaussian blur function from the editing menu on the Final Cut Pro program. The level of blur range from 0 to 100, with a value of 0 representing no blur to 100 representing total blur. The initial value that was used was a value of 13. Once this blur level was applied, the overall degradation effect was deemed sufficient by the experimenter. The blurred image of Chris Tarrant is provided in Figure 5.b below. However, it is difficult to fully appreciate the effectiveness of the blur filter as the image here is in a still format, whereas participants were tested with moving images.

Figure 5.b: Image of Chris Tarrant with blur filter at level 13 applied using Final Cut Pro.

*Negative Faces.*

Each image was firstly converted into monochrome through Final Cut Pro. Once this was completed, each image was transformed into negative through the image effects menu on Final Cut Pro. The photographic negative of Chris Tarrant is provided in Figure 5.c.

Figure 5.c: Image of Chris Tarrant converted into photographic negative using Final Cut Pro.



After each clip (famous and non-famous) had been modified using the three procedures outlined above, each version (inverted, blurred and negative) was transported as a QuickTime file for presentation. The presentation order of the clips across the three degradation conditions was in a fixed random order. The final size of each image measure approximately 14 inches (35cm) along the diagonal.

### Design.

This pilot study involved a one factor between subjects design. The single factor of facial degradation included three levels, which were inverted faces, blurred faces and negative faces. Ten participants were randomly allocated into one of the degradation groups. The dependent measures taken were the same as the ones used in Chapter 4. These measures centered around the performance measures presented in table 4.1.a (see page 145) and included the proportion of hits and false alarms in each condition, correct recall of the person's occupation and correct recall of the person's name as conditionalized on recognition. Participants' correct and incorrect responses at each level of confidence were used to generate ROCs and z-ROCs for each condition.

### Procedure.

Participants were tested in three groups of ten participants per group. Each participant within each group was seated in front of an Apple Mac computer and given a response book. Testing instructions and responding was the same as in Chapter 3 (see page 144).

Testing took the form of two experimental blocks with 48 trials (24 famous and 24 non-famous) within each block. Half the participants completed Block 1 first, followed by Block 2, whilst half the participants did Block 2 first, followed by Block 1. The items were presented in a fixed random order.

Participants were seated 30cm in front of the computer monitor. They were instructed that they would be presented with altered facial images, some of which belonged to famous people. They were specifically instructed to remain absolutely still and to refrain from any head-movements which may be triggered to compensate for the presentation of the images (i.e; the inverted images).

The participants were instructed to decide for each extract, whether the item was famous or non-famous. If they thought that the item was famous, they had to indicate OLD and if they thought the item was non-famous they had to respond by indicating NEW on their response sheet. They also had to provide a confidence rating for each OLD/NEW decision on the basis of Yonelinas et al's (1999) 6-point confidence scale (6=sure OLD, 5=probably OLD, 4=maybe OLD, 3=maybe NEW, 2=probably NEW, 1=sure NEW ).

If they made an OLD judgment, participants were required to give a description of the person's occupation, along with any other biographical details (eg, if they are associated with anybody famous, etc). They were instructed not to guess the person's occupation. Finally, they were also required to name the person. They were required to write each response in their response booklet.

The presentation of each clip was at the participants' control, however they were prompted to work as quickly as possible by the experimenter and to press the space bar to move onto the next clip. Therefore the main difference between this study and the one carried out in Chapter 4 was that the presentation rate in this pilot study was at the participants' control and the participants themselves were required to write down their responses. This change in procedure arose mainly because of practical and time limitations, which also meant that participants were tested in groups rather than individually.

After the first block was completed, participants were allowed a short break before starting the second block of extracts. Once the experiment was completed, participants were de-briefed and given a book voucher for their participation. As part of the de-briefing session, each participant was given a list of names of all the celebrities that they saw. They were asked to place a tick next to the name if they were familiar with that celebrity. This task was done to ensure that all the celebrities were well known to the participants. These responses were not used for statistical purposes.

## Results.

The results section is divided into two parts. The first part focuses on the performance measures across the three degradation conditions. The mean values for each of the performance measures across the three degradation techniques were analysed by a series of one-sample t-tests using the values from the voice condition obtained in Chapter 4 as test values. For completeness, the means from the standard face condition (Chapter 4) are also reported, however they were not used in any of the analyses. These values are presented in table 5.1.a below, with the values in bold format representing the values which do **not** statistically differ ($p > 0.05$) from the voice condition. The same strict marking criteria were applied when assessing participants' occupational descriptions (see page 75 for details). The second part of the results section presents the ROCs and z-ROCs for the three conditions. These are assessed in terms of the contributions of recollection and familiarity.

Table 5.1.a: Performance measures for inverted, blurred and negative faces. Standard faces and voices means are taken from Chapter 4. Standard deviations in parentheses. Values represent proportion correct.

| | Faces Chapter 4 | Voices Chapter 4 | Inverted Faces | Blurred Faces | Negative Faces |
|---|---|---|---|---|---|
| Hits | 0.92<br>(0.05) | 0.72<br>(0.13) | 0.84<br>(0.09) | **0.73**<br>**(0.08)** | 0.68<br>(0.06) |
| False Alarms | 0.14<br>(0.12) | 0.28<br>(0.13) | **0.29**<br>**(0.15)** | **0.28**<br>**(0.21)** | **0.28**<br>**(0.14)** |
| A' | 0.94<br>(0.04) | 0.81<br>(0.06) | **0.85**<br>**(0.08)** | **0.79**<br>**(0.13)** | **0.78**<br>**(0.07)** |

| | Faces Chapter 4 | Voices Chapter 4 | Inverted Faces | Blurred Faces | Negative Faces |
|---|---|---|---|---|---|
| B D | -0.17<br>(0.67) | 0.04<br>(0.49) | **-0.22**<br>**(0.53)** | **0.07**<br>**(0.44)** | **0.12**<br>**(0.43)** |
| Occupations recalled conditionalized on recognition | 0.83<br>(0.12) | 0.54<br>(0.32) | **0.54**<br>**(0.32)** | **0.62**<br>**(0.25)** | **0.67**<br>**(0.26)** |
| Names recalled conditionalized on recognition | 0.72<br>(0.19) | 0.50<br>(0.32) | **0.50**<br>**(0.32)** | **0.58**<br>**(0.28)** | 0.59<br>(0.23) |

It is interesting to note that all three degradation techniques were successful in removing the ceiling effects in the standard face condition. Comparing the three techniques, it appears that the blurred condition elicited performance that was most closely matched to the voice condition. None of the performance measures for blurred faces differed significantly from the voice condition. Therefore, blurred faces were matched in terms of overall discriminability, bias and knowledge of famous items as seen in the occupational and name measures.

The inverted condition, produced only one measure that differed significantly from voices. This was the mean hit rate, which was significantly higher for inverted faces than voices. On all of the other measures, performance with inverted faces did not differ from voices. Negative faces produced a hit value which was significantly lower than the voice condition. In addition, the name recall data significantly differed from the voice condition, so that there were significantly more names recalled to negative faces than voices. This means that the negative faces were not matched with voices in terms of overall knowledge of famous items.

### *Inverted Faces - ROCs and z-ROCs .*

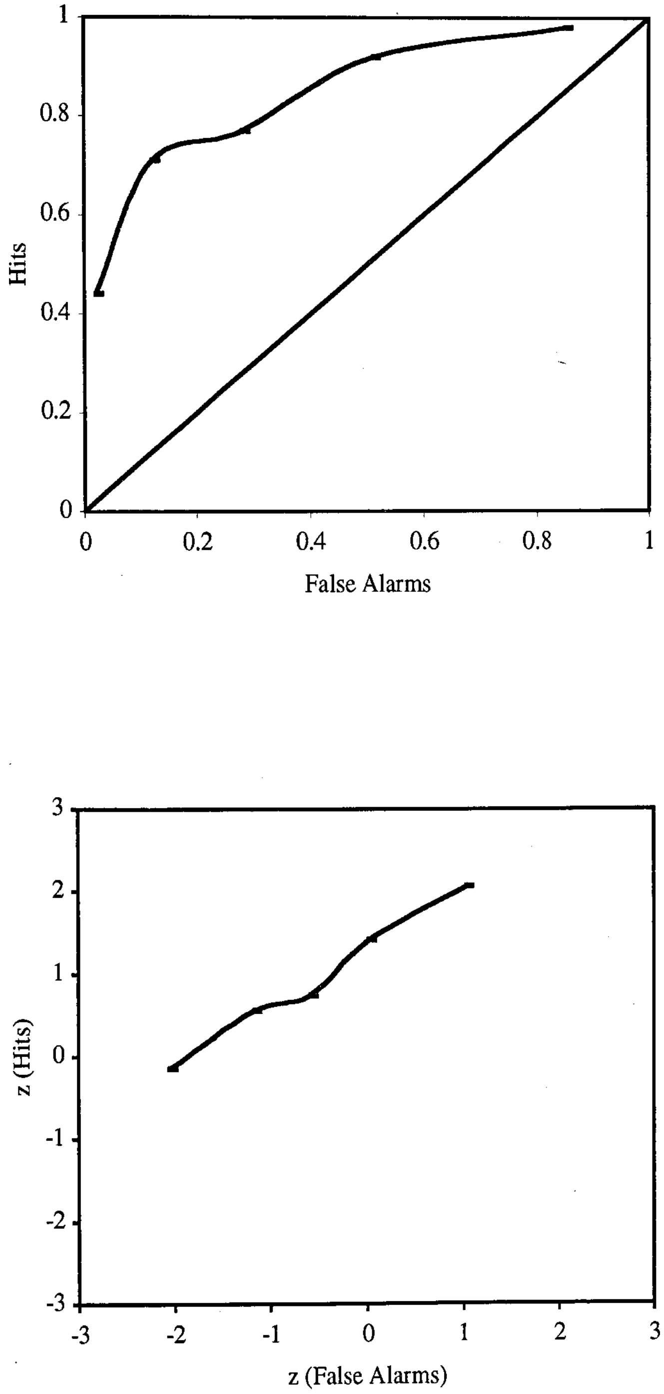
Receiver Operating Characteristics (ROCs) were generated from the cumulative probabilities of hits (x axis) and false alarms (y axis) across the different confidence levels. The ROCs and its corresponding z-ROCs for inverted faces are displayed in figure 5.1.a.

The decrease in accuracy can be viewed in this figure in terms of the ROC moving towards the diagonal line (no discrimination). This figure can be compared with the standard face ROC in Chapter 4 (page 154), which demonstrates how an ROC is pushed away from the diagonal as performance approaches ceiling. By presenting faces in an inverted format alters the position of the original ROC curve, but how does inversion affect the shape of the curve?

A close inspection of figure 5.1.a (top panel) shows a slight distortion at around the middle confidence values of the ROC curve which makes it difficult to assess the symmetry of the curve. The slight distortion that was present in the ROC (top panel) has also transferred to the z-ROC (bottom panel). Although the z-ROC for inverted faces did show some curvature, this deviation from linearity is not consistent with Yonelinas' (1999) dual-component model. As a consequence, only a standard linear regression was applied to the z-ROC. This produced the following equation, z(hits) = 0.7132(zfalse alarms) + 1.3123, $R^2 = 0.987$. As the slope (0.71) was not equal to 1.0 (CI = 0.56 – 0.86), the equal variance familiarity model was ruled out. It therefore appears that the pattern of performance exhibited by the inverted face condition is most adequately captured by an unequal variance familiarity model.

Some additional comments that are worth making with regards to the ROC and z-ROC is that there appears to be a third parameter contributing to performance, one which is not captured by recollection and familiarity. This is reflected by the misplaced 'dip' in both curves. Such an observation was found in the previous chapter for voices (see page 155) and it was concluded on the basis of previous research that such deviations in z-ROC points may represent a guessing component contributing to recognition (Rotello et al, 2000).

Figure 5.1.a: Average ROCs (top panel) and average z-ROCs (bottom panel) for the inverted face condition. Diagonal line represents no discrimination.

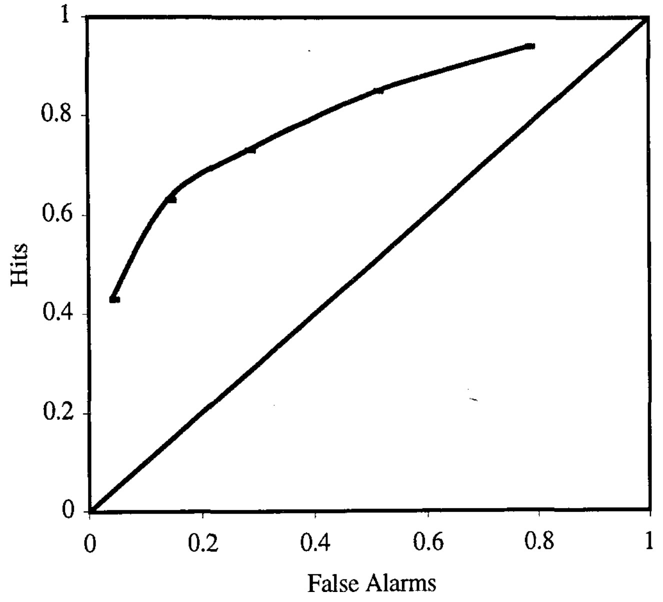


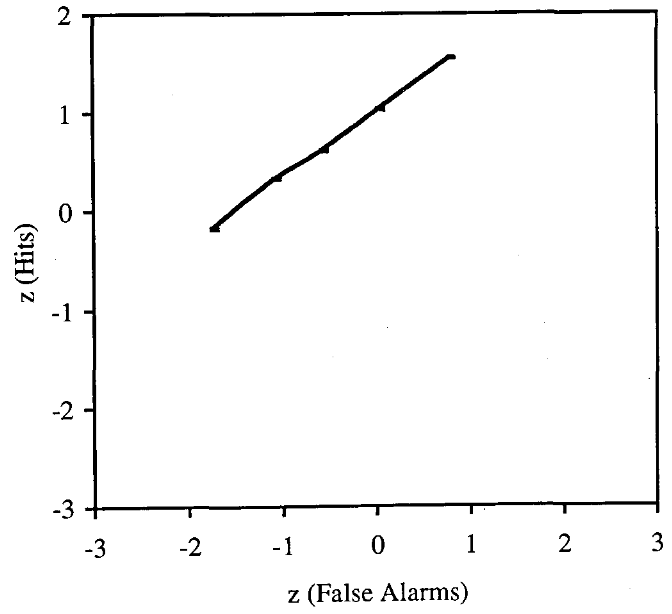
*Blurred Faces – ROCs and z-ROCs.*

The ROC and z-ROC for the blurred face condition is displayed in figure 5.1.b. Again it is clear, that altering the visual image, in this case by blurring, relocates the ROC towards the diagonal (top panel). The ROC generated by blurred faces is curved and asymmetrical along the diagonal. The corresponding z-ROC is predominantly linear. This particular curve confines the analyses to the equal variance familiarity model and the unequal variance familiarity model. The z-ROC generated by blurred faces was examined to see which of these two models is more appropriate in describing the data.

The blurred condition also appears to exhibit a slight 'dip' in the middle of the curve, however this is not as marked as with the inverted face condition. The analysis of the z-ROC was carried through a standard linear regression analysis. This yielded the following equation, z(hits) = 0.6785 (zfalse alarms) + 1.027, $R^2 = 0.9988$. Since the slope (0.68) was not equal to 1 (CI = 0.63 – 0.72) the equal variance familiarity model was ruled out and the unequal variance familiarity model was accepted as the most appropriate model for the data.

Figure 5.1.b: Average ROCs (top panel) and average z-ROCs (bottom panel) for the blurred face condition. Diagonal line represents no discrimination.





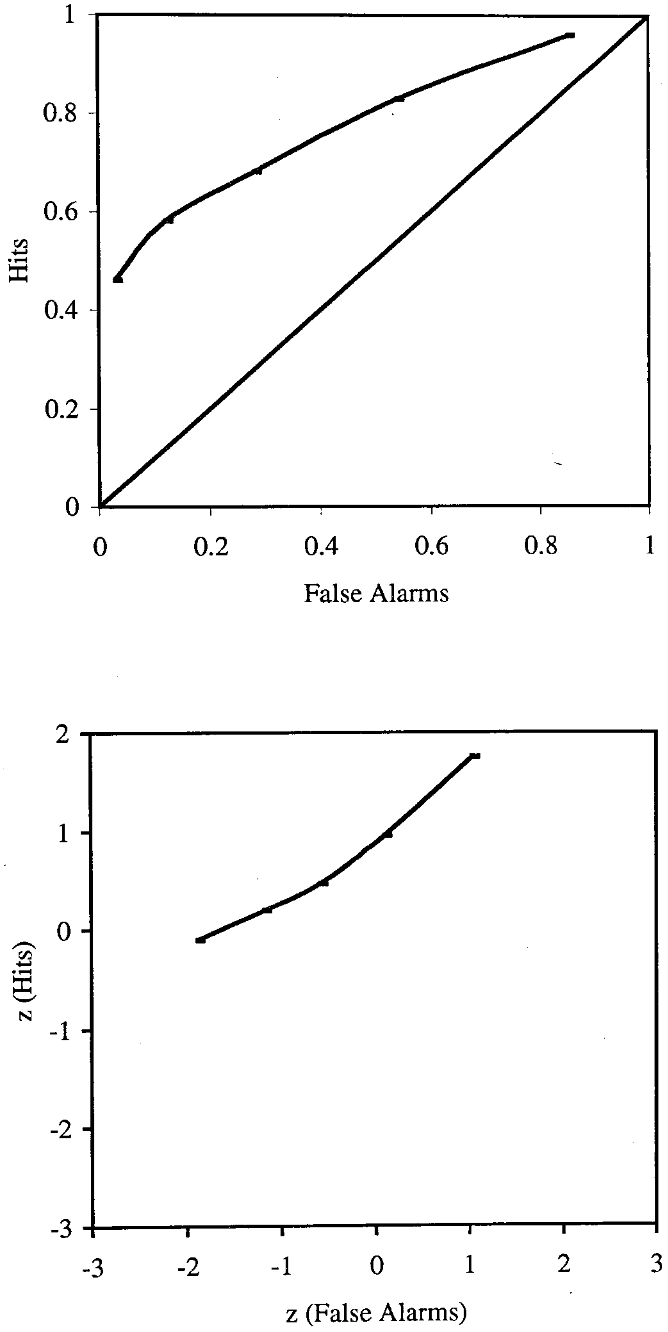
*Negative Faces – ROCs and z-ROCs.*

The final set of ROCs and z-ROCs were applied to the data from the negative face condition. The ROC and z-ROC are displayed in figure 5.1.c. Presenting faces as photographic negatives also has the effect of shifting the ROC towards the diagonal (top panel). However, the shape of the curve in the top panel of figure 5.1.c does show some evidence of a linear component. The linearity of the model is corroborated further by a curved z-ROC. However, the ROC is not completely linear, this can be seen by the front end of the ROC showing some elements of curvature. With this in mind, it would therefore appear that the ROC and z-ROC generated by the negative face condition closely supports the dual-process model as outlined by Yonelinas (see figure 4.4a). If this is the case, then the regression with the quadratic function should account for more of the variance than a standard linear regression. Both methods were applied to z-ROCs. A quadratic function was deemed suitable for the negative face z-ROC as the curvature is located at the front end of the z-ROC - a position that is consistent with z-ROC analyses (Yonelinas, 1999).

The standard linear regression for the negative face condition produced the following equation, z (hits) = 0.6341(zfalse alarms) + 0.9721. This equation accounted for 97.4% of the variance. When a quadratic function was incorporated into the analysis, the following equation was produced, z(hits) = 0.1128(zfalse alarms)$^2$ + 0.7268(zfalse alarms) + 0.8754. With this additional component, the variance accounted for now increased to 99.9%. Critically, the statistical analysis conducted on the data showed that the variance accounted for by this additional component is significant, $F_{INC}(1, 2) = 89.13$, $p < 0.05$.

The pattern of analysis from the ROCs and z-ROCs therefore seems to indicate that a recollection process along with a familiarity process may be contributing to recognition performance in negative faces. This is reflected by a linear component in the ROC and supported further by a curved z-ROC. This is a particularly interesting observation, given that the overall level of recognition is lower than standard faces (table 5.1.a).

Figure 5.1.c: Average ROCs (top panel) and average z-ROCs (bottom panel) for the negative face condition. Diagonal line represents no discrimination.



## General Discussion.

One of the central aims of this chapter was to apply various image degradation techniques to the selection of famous and non-famous faces used throughout this thesis in order to produce a face modality condition that was comparable with the performance measures obtained for voices (table 5.1a).

Each of the three degradation techniques were successful in removing the ceiling effects observed with faces under normal presentation conditions. Starting with the inverted face condition, the hit and A' values faired reasonably well with earlier studies on inversion effects. For example, Lander et al (1999) achieved a mean A' value of 0.76 for their inverted-moving face condition. This value is somewhat lower than the A' value obtained in the current study, however this difference should be seen in the context of the different presentation rates across the different studies. Specifically, the lower A' rate in Lander et al's (1999) study may have come about because the stimulus display time was smaller (2.5 seconds) in their study, compared with this study (7 seconds). As inversion requires the participant to mentally rotate the image to 'correct' it (Collishaw and Hole, 2000), the longer duration employed in the current study may have provided participants with additional time to carry out this process. This in turn could have resulted in the easier discriminability of famous over non-famous items.

In terms of how the inverted face condition compared with the voice condition, the series of t-tests only produced one performance measure that differed significantly from the voice condition and this difference was found in the hit rates. The hit rate for the inverted face condition was found to be significantly higher than in the voice condition. As a result, this degradation technique was not deemed suitable enough as a comparison condition for voices.

Presenting faces in a blurred format also proved to be an effective technique in reducing overall recognition. However, it is rather difficult to compare the blurred values from this pilot study with those obtained by previous researchers (Lander et al, 2001, Hanley and Turner, 2000). This is mainly because the two studies used different methods of applying the blur effect. For example, Hanley and Turner (2000) applied the blur effect by altering the focus lens on the projector. Lander et al (2001) on the other hand, digitally applied the level of blur through the use of computer software. Although both of these methods can efficiently reduce overall face recognition, this pilot study favoured the digital method for practical purposes and also to maintain the facial size ratio used throughout this thesis.

The critical aspect of the blurred faces data relates to how effective this technique is in providing a testing condition that is comparable to voices. The series of t-tests carried out did not produce any significant differences between the blurred faces and voices for any of the performance measures. This suggests that presenting faces in a blurred format is an excellent way to not only remove ceiling effects in face recognition but also to provide a suitable matched condition for comparisons with voice recognition (table 5.1.a).

The final degradation technique that was employed in the current pilot study involved presenting the faces as photographic negatives. This particular degradation technique produced a hit rate of 0.68 which was the lowest of all three degradation techniques. Furthermore, the series of t-tests carried out revealed that this hit rate was also significantly lower than the hit rate obtained for voices.

In relation to previous research that has used negative faces, Knight and Johnston (1997) obtained a percentage correct value of 61.4%, which is reasonably similar to the hit rate obtained in this study (table 5.1.a). Looking at performance beyond hit and false alarms rates, the name data is highly comparable to the values obtained by Lander et al (1999). The study by Lander et al (1999) produced a mean recall rate of 57% for moving negative faces – a value which is almost identical to the name recall data for negative faces in this study. Moreover, this particular performance measure was found to be significantly higher than the name recall measure for voices (table 5.1.a). As the negative face condition produced two measures that differed significantly from the voice condition, this particular degradation technique was not considered to be an appropriate condition to compare with voice recognition.

One of the problems encountered in the analyses of ROCs and z-ROCs for the face condition in the previous chapter was that the ceiling effect in the face condition made it difficult to argue for a recall component in face recognition. Related to this issue of accuracy and recollection, Yonelinas, Kroll, Dobbins, Lazzara and Knight (1998) have shown that high levels of accuracy are not necessarily related to the occurrence of recollective processes. This means that even when recognition performance was made harder, recollection continued to contribute to recognition performance.

The current study aimed to remove the ceiling effects by applying degradation techniques to standard faces. As all techniques proved successful (see table 5.1.a), the next stage of the analysis looked at how this reduction in overall performance affected the ROCs and z-ROCs generated by each condition. Starting with the inverted condition, the ROCs that was produced appeared curved, however the symmetry of the curve was difficult to assess due to a slight distortion at the middle range of the curve. When the corresponding z-ROC was analyzed this also exhibited an unconventional 'dip' in the middle range. However, in spite of this slight distortion, the z-ROC was deemed sufficiently linear for standard linear regression to be applied to it. The conclusions made from these analyses were that inverted faces are captured by an unequal variance familiarity model. In other words, recognizing items under this degradation condition is likely to be based on assessments of familiarity with each item, with the most familiar item being accepted.

This unequal variance familiarity model was also used to account for the data produced by the blurred faces (see figure 5.1.b). This degradation technique also produced a slight 'dip' in the z-ROC. These 'dips' in z-ROCs have been observed in the previous chapter with the voice condition. A review of the ROCs literature suggested that this dip can be attributed to a type of guessing component which operates across the middle confidence categories (Rotello et al, 2000).

A different conclusion is drawn for the negative face condition. The overall hit rate is almost 25% less than the standard face condition. Yet, the ROCs and z-ROCs produced by this condition shows evidence of a possible recall component contributing to recognition. This is reflected by a rise in over 2% of the variance when the z-ROC was analyzed with a quadratic function. This observation is consistent with Yonelinas et al's (1998) study. In other words, recall is likely to contribute to performance (along with familiarity), even when overall accuracy decreases. This particular finding suggests that a recollective process may indeed be contributing to face recognition.

What remains unaccounted for however, is why the negative condition produced a recall component but inverted and blurred faces did not? One tentative explanation may be found in the way in which inverted and blurred faces are processed. As outlined in the introduction, successful face recognition is likely to entail the complimentary contribution of two qualitatively different cognitive processes. One of the processes deals with the configurational properties of a face, whereas the other one processes the featural properties of a face (Collishaw and Hole, 2000). If one route is inaccessible, then recognition can still continue via the other route. This configural / featural distinction was supported by Collishaw and Hole's (2000) degradation condition which inhibited the use of both routes by presenting faces that were blurred and inverted simultaneously. Under this condition, recognition was considerably impaired relative to conditions where the two degradation techniques were applied separately.

With this configural / featural distinction hypothesis in mind, it may be the case that both types of processing need to be accessible in order for recollection to contribute to performance. The main properties of a face that are impaired in a negative face are its depth cues. However, as pointed out by Knight and Johnston (1997) these depth cues are reinstated through movement. Therefore a moving negative face may have both configural and featural properties available, which in turn may explain why recollection is present in this technique and not in the other two.

However, this idea is only speculative and is based on the types of processing that are promoted / inhibited in inverted and blurred faces (Collishaw and Hole, 2000). Additional research is required to examine the type of processing that is required for negative faces and whether presenting such images in a static format can differentially affect the contributions of recollection and familiarity.

To conclude, the three degradation techniques that have been used in this study have all removed the ceiling effects found for faces in Chapter 4. Across the three conditions, the blurred face condition proved to be the most effective condition for comparisons with the voice condition. Not only did the blurred faces elicit similar hit, false alarm, discriminability and bias values, but blurred faces were also matched with voices in terms of overall knowledge (occupations and names). Additionally, the blurred face condition was chosen as the 'new' face condition as it represents the most realistic form of degradation (Lander et al, 1999).

The findings from this pilot study will form the foundations of the final experimental chapter of this thesis. Specifically, participants' recognition memory will be tested using the standard presentation of faces, voices and now blurred faces as a control for the voice condition. The role of two qualitatively different memory processes shall move away from the ROC / z-ROC procedure as adopted by Yonelinas (1999). Instead, recollection and familiarity will be looked at from the perspective of Tulving (1985), whose research subsequently lead to the Remember / Know (R / K) procedure. This approach assesses the qualitative aspects of recognition memory as directly reported by participants. Recent modifications to the R / K procedure have seen the addition of a guess category incorporated alongside the original two responses (Gardiner, Ramponi and Richardson-Klavehn, 2002). Considering the problems associated with the ROCs and z-ROCs procedure (Heathcote 2003, Rotello et al, 2000), this new category will hopefully lead to a clearer picture of the roles of recollection and familiarity to familiar face and voice recognition, especially now that these two processes will be assessed when overall recognition between the two modalities is matched through the use of blurred faces.

# Chapter 6: The R and K paradigm. Can it be applied to the recognition of familiar faces and voices?

## 6.1: Introduction.

Chapter 4 reported one approach to understanding the differences between face and voice recognition by analyzing the ROCs and z-ROCs produced by each modality. The experiments reported in this chapter extends this line of research by investigating the relative contributions of remembering and knowing in person recognition memory.

Chapter 4 introduced Mandler's (1980) dual component theory of recognition memory. This theory was of seminal importance in the psychological literature as it led to the subsequent development of the pioneering Remember and Know (R/K) procedure by Tulving (1985) and Gardiner (1988). This procedure allowed psychologists to capture these distinct memory processes as originally put forward by Mandler (1980).

Mandler's 'butcher in the bus' analogy was used to illustrate how the recognition of a familiar person may come about by recognizing the face as one that has been encountered before (familiarity mode) or by retrieving the relevant context (elaborative processing / recollection mode). In this particular instance, the relevant context would be the supermarket, which would be used to produce enough cue specificity to retrieve the memory, in this case, the butcher from the supermarket (Mandler, 1980, pp 252-253). The critical question that is addressed in chapter 6 is whether R responses will predominate in face recognition and K responses will predominate in voice recognition.

## 6.2: The R and K paradigm.

The R/K paradigm was based on Tulving's (1985) idea that these two memory processes elicit qualitatively different subjective states of awareness and the paradigm itself has been used in conjunction with the standard yes/no recognition procedure. In a standard recognition experiment, a participant may recognize a previously encountered item because it brings to mind a specific experience in which the item was encoded. This form of retrieval falls under Mandler's (1980) inter-event elaborative process and was subsequently re-named by Tulving as recollection. Other terms associated with this mode of recognition include remembering.

Recognition of a previously encountered item can also be achieved without mentally re-experiencing aspects of the encoding episode. In such instances, recognition of the item is achieved through strong feelings of familiarity attributed to the item as initially proposed by Mandler (1980). Tulving continued to use Mandler's (1980) definition for this process, however researchers within this branch of memory tend to refer to this process as knowing.

Tulving (1985) argued that recollection and familiarity represent retrieval from two distinct, yet partially overlapping memory systems. Conscious recollection (remembering) is thought to be a product of the episodic memory system whereas feelings of familiarity (knowing) represents retrieval from the semantic memory system. This thesis will use the terms remembering and knowing to refer to Tulving's earlier terms of recollection and familiarity, respectively.

In a typical recognition memory experiment that uses the remember (R) and know (K) paradigm, participants are given a selection of materials to study, for example a list of words. In the test phase of the experiment, participants are presented with words from the original list and some new words that they had not encountered in the study phase. For each word, participants have to decide whether they recognize the word or not. For the items that they recognize, participants are asked to classify their recognition into one of two categories, namely, 'remember' (R) and 'know' (K). Participants are often instructed that an R response means that recognition is accompanied by a specific recollection of the item's occurrence, such as an association or an image that the item evoked. Participants are required to give a K response when they are certain that they had encountered the item previously, but the item brings nothing to mind other than a feeling of familiarity. What makes the R/K procedure a useful tool in recognition memory is that it can allow the researcher to determine the type of recognition that has occurred. Objective measures such as proportion correct, may be equivalent in two different experimental conditions, yet the contribution of remembering and knowing may differ.

Following Tulving's (1985) seminal work, researchers have established a variety of variables that produce systematic dissociations between remember and know responses. For example, Gardiner and Java (1990) showed that whilst there was very little effect of word type on overall recognition, when recognition was broken down into remembering and knowing, words were associated more with remembering than knowing. Non-words on the other hand were associated more with knowing than remembering.

Experimental manipulations are not merely restricted to word stimuli. Interesting observations have also been found with a variety of other stimuli, including Polish folk songs (Gardiner, Kaminska, Dixon and Java, 1996). This latter study has found an interesting pattern of results whereby variable manipulation had parallel effects on remembering and knowing.

Evidence for functional dissociations between remembering and knowing have also been observed for different subject populations as well as for different experimental manipulations. For example, Perfect, Williams and Anderton-Brown (1995) compared word recognition between older and younger adults. The initial results showed that the overall recognition between the two groups did not differ. However, there was an apparent difference in the nature of the recognition reported. The younger group reported many more remember responses than the older group, whereas the older group reported more know responses than the younger group.

An alternative way to assess the relative contributions of remembering and knowing is to examine the contents of these subjective states of awareness. One such study was carried out by Gardiner, Ramponi and Richardson-Klavehn (1998). The initial experimental stage included a standard yes/no distinction to previously studied words and new words, followed by remember, know or guess responses for each item that was recognized. The addition of the guess response will be discussed in more detail in the following sections, but for current purposes, the guess response was included so that participants were not forced to use know responses if they could not remember the item's previous occurrence. After the participants made their decision, the researchers randomly selected some items to which the participants gave a remember, know or guess response. In order to reduce the opportunity for confabulation, participants were asked to justify their recognition decision only and not their reported states of awareness.

In reviewing the descriptions given, Gardiner et al (1998) were also able to examine whether participants were using remember, know and guess responses in a valid way. The general finding with remember responses was that there was strong contextual association with the studied word. These contextual associations were broken down into 5 sub-categories by the authors. For example the extra-list association category tended to produce associations from outside of the study phase, so that the word 'kilt' was recognized by one participant because they remembered thinking of a Scottish man. Another category devised by the authors incorporated the word's imagery features. In one particular instance, a participant who remembered the word 'gun', did so because they could remember the image of a gun.

Moving on to the descriptions for know responses, the authors noted that whilst participants were aware of encountering the studied word, they could not pin-point the exact context of the encounter. Most of the participants' descriptions for know responses were along the lines of 'familiar only', 'rings a bell'. However, it is worth to note that the descriptions were not completely absent of contextual information, but merely incomplete or vague. For example the word 'kite' triggered one participant to describe it as "something to do with the sky" (Gardiner et al, 1998, appendix 1). Therefore, it appears that know responses lack the elaborate contextual detail that is associated with remembering.

Guess responses on the other hand, were just as likely to be associated with studied items as with non-studied items. Guess descriptions tended to be associated with low levels of confidence and related to judgmental strategies rather than memory for the studied word.

This study by Gardiner et al (1998) provides an elegant database of how remembering, knowing and guessing vary in terms of contextual information. It is possible that confabulations may occur in remembering and knowing (Dalla Barba, 1993). In light of this, the experiential approach adopted by Gardiner et al provides a useful way to assess the validity of participants' responses.

Although the studies presented so far have demonstrated how remembering and knowing can be systematically dissociated across different experimental variables and different populations, the theoretical explanations of these dissociations are open to debate.

Some theoretical explanations put forward since Tulving's episodic/semantic framework includes the processing account proposed by Rajaram (1996). According to this theory remember responses are largely governed by conceptual variables such as level of processing whereas know responses are governed by perceptual variables such as same versus different study test modality. This processing account has been extended to accommodate the role of distinctiveness, which some researchers argue is the crucial factor that underpins remembering (Mäntylä, 1997). Even though there may be differences in how these theories attempt to explain remembering and knowing, taken together the research presented so far provides support for the idea that recognition memory is governed by two distinct components, however those components are conceptualized.

### 6.3: Remembering and Knowing — a quantitative explanation?

The dual-component view of recognition memory covered in the previous section has recently come under attack. Whilst researchers such as Mandler (1980) and Tulving (1985) advocate that differences in recognition memory rest at a qualitative level, the arguments developed by Donaldson (1996) pose a major challenge to this approach. He argues that remember and know responses merely reflect quantitatively different responses to the same memory trace.

Donaldson proposes that the dissociations observed by Gardiner and others can be easily generated from a single strength model of recognition memory.

Donaldson summarizes remembering and knowing by a means in which "participants simply divide their yes responses into strong remember responses and weak know ones ... remember responses represent nothing more than conservative yes responses" (pp, 524). In other words, overall recognition, that is, remembering plus knowing, reflects more lenient response criteria. Remember only responses, on the other hand, reflect more stringent criteria.

Donaldson used signal detection modeling to test these assumptions of the unitary strength model. For the single strength model to be upheld, memory strength, as measured by A' prime, should be the same whether calculated from remember plus know hits and false alarms or remember only hits and false alarms.

For overall recognition (R plus K) Donaldson obtained an A' value of 0.825, compared to 0.837, for remember only responses. The difference between these two values was not significant. This result was used by Donaldson to strengthen the argument that know responses do *not* reflect an additional source of memory.

Although Donaldson's work has demonstrated how signal detection modeling can be used to explain remembering and knowing, and in turn provide support for the single strength model, a recent modification of the R/K procedure has re-ignited support for the dual component model.

In a recent meta-analytic study by Gardiner, Ramponi and Richardson-Klavehn (2002), they obtained A' values of 0.813 for remember plus know responses compared to 0.787 for remember responses alone. The difference between these two values was significant. This meta-analysis differs from Donaldson's study in that an additional response - a guess response - was included in the analysis. Gardiner et al argue that this new addition to the R/K paradigm, means that participants are no longer forced to give a know response if they are not able to remember. Thus, know responses are uncontaminated if guess responses are reported. A guess response is defined as having some other reason to believe that the item was encountered previously, some other reason than remembering and knowing. When guess responses are added to remember and know responses, memory strength is decreased. More importantly, guess responses show no memory for studied and non-studied items, with typical A' prime values being around 0.50. On the other hand, A' prime values for remember responses and know responses were above chance.

This meta-analytic study by Gardiner et al (2002) demonstrated that when know responses are extracted from guessing, then knowing does indeed provide an additional source of memory. Such findings add converging evidence in support of the dual component model.

Replacing remember and know judgments with confidence ratings provides another means to test the single strength model. If remember and know responses merely reflect very strong and very weak memory traces respectively, then these states of awareness should map onto confidence judgments. Such a hypothesis has been tested by a number of researchers who in turn have collectively shown that remember and know responses are not equivalent to confidence ratings. This finding occurs whether the confidence scale only has two levels, such as sure/unsure (Gardiner and Java, 1990), or whether the confidence scale includes four ratings (Manytla, 1997). Furthermore, this finding is maintained even when the same participants are asked to make both R/K and confidence responses (Rajaram, Hamilton and Bolton, 2002). More importantly, Rajaram et al's findings showed that amnesic patients compared to their matched controls and college students did not show the typical interaction in subjective awareness for words over non-words. The crossover effect in word versus non-word recognition shows that remembering is more likely to be associated with words, whereas knowing is more likely to be associated with non-words (Gardiner and Java, 1990). The fascinating aspect of these results was that whilst the amnesic group differed in their subjective awareness for words and non-words, they did not differ from the control group in terms of their confidence judgments. The amnesic group, like their controls, made more sure than unsure judgments to both words and non-words. These results from Rajaram et al's (2002) study shows that the dissociations between remembering and knowing and confidence can also occur at a neuropsychological level.

Since Tulving's (1985) influential work and the subsequent development of the remember and know (R/K) paradigm (Gardiner, 1988), a number of studies, some of which have been reviewed here, have examined variables that produce dissociations between remember and know responses. Whilst these studies have used these dissociations to support the view that remembering and knowing entails two distinct memory components or systems, others view remembering and knowing to represent two different decisions to the same memory trace.

Studies that have used the R/K paradigm have used a variety of test materials ranging from words and pictures to folk music, however only a handful of studies have incorporated the R/K paradigm to the understanding of person recognition. This is an extremely surprising state of affairs considering that R/K is consistently defined within a person recognition context (Gardiner et al, 2002 and Perfect et al, 1995).

In this chapter, the distribution of R, K and Guess (G) responses will be examined across faces and voices. The defining feature of this chapter is the inclusion of an additional face condition to serve as a matched control for voices. The analyses from the previous chapter (Chapter 5) illustrated that blurring faces provides an adequate match for voices in terms of overall recognition performance. Therefore the critical area of interest will be in terms of the distribution of R and K responses when overall recognition between faces and voices is matched (blurred faces vs voices). Before the hypotheses are discussed in detail, the following section provides an overview of some of the studies that have incorporated the R/K paradigm in person recognition memory research.

### 6.4: The R/K paradigm and person recognition.

Mandler's (1980) butcher in the bus analogy is perhaps one of the earliest examples of how remembering and knowing can contribute to person recognition memory research. Twenty years on, the R / K distinction continues to be defined in person recognition terms.

For example, Gardiner et al (1998, Appendix 2) define remembering in terms of 'recognizing someone's face and remembering other specific events about that person, such as talking to them'. Perfect et al (1995, pp 173) define knowing in terms of 'recognizing a face in the street, but not being able to place them or remember anything about them other than finding the face familiar'. Despite remembering and knowing being defined in terms of person recognition, the application of the R/K paradigm to this area of memory has been very sparse indeed.

The few studies that have incorporated the R/K paradigm to the study of person recognition include the work done by Mäntylä (1997) and more recently Brandt, Macrae, Schloerscheidt and Milne (2003). Brandt et al have outlined a number of different issues that can be addressed through the use of the R/K paradigm. One issue relates to the initial recognition experience; whether this is achieved by remembering or by knowing the previous personal encounter. An additional issue that is also raised is whether the types of subjective awareness vary in relation to the social cues that are presented to participants. As it has been argued throughout this thesis, the ability to recognize others is not necessarily limited to face recognition, other modalities, such as the person's voice also provide an alternative means to person recognition (see figure 1.6b by Belin et al, 2004, page 25). This particular social cue of person recognition will be discussed in the following sections, for now the main conclusions from the work by Mäntylä (1997) and Brandt et al (2003) shall be presented.

Mäntylä's (1997) study looked at the effects of encoding on subsequent reports of remembering, knowing and guessing. Participants in this study were given two encoding conditions. In the distinctive condition participants were instructed to rate each face in terms of its distinctiveness. In the relational condition, participant's had to rate the similarities between faces by sorting them into social categories such as 'party-goer' and 'sporty type'.

The initial results showed that the overall recognition between the two groups was almost identical. However, when the results were analysed in terms of the proportion of remember and know responses, the distinctive encoding condition produced more remember responses than the relational condition. The relational condition, on the other hand produced more know responses than the distinctive condition. In the same paper, Mäntylä (1997) also investigated the effects of changes in facial appearance on subsequent recognition performance. The results showed that changes in appearance produced more recognition based on knowing rather than remembering. This latter finding has a direct application to eyewitness memory by showing that a witness may not necessarily have an explicit recollection (remembering) of the suspect, especially if there are changes in appearance from the original encounter.

The study by Brandt et al (2003) looked at the role of facial distinctiveness on recognition performance. Their main findings showed that remembering was more strongly associated with distinctive than typical faces, whereas the effect was reversed for know responses.

Whilst these studies by Mäntylä (1997) and Brandt et al (2003) have demonstrated how the R/K paradigm can be effectively applied to the recognition of novel faces, what now remains to be addressed is how the R/K paradigm can be used to study the processes involved in familiar face and voice recognition. The emphasis in this chapter is the recognition of voices compared to faces and how recognition may differ in terms of subjective awareness for these two modalities.

### 6.5: Rationale.

Comparing face and voice recognition performance has been investigated by Hanley, Smith and Hadfield (1998) and re-addressed in Chapter 3 of this thesis. For both studies, faces were associated with higher hit rates and lower false alarm rates than voices. However, the interesting finding of these studies was found in the occupational recall performance data. The five experiments reported in Chapter 3 showed that faces and voices are different in the extent to which they elicit person specific information. In terms of person recognition breakdown, there were significantly more instances in which a voice was associated with a familiar-only error than a face. This finding persisted even when level of familiarity was held constant across the two modalities. This suggests, by definition of a 'know' response (Gardiner et al, 1998), that voice recognition relative face recognition may be based on knowing. This is an appropriate inference to make given the study by Maylor (1995) who used participants' lack of descriptions for television theme tunes as an index of knowing. The studies carried out by Hanley et al (1998) and in Chapter 3 did not use the R/K paradigm. Therefore this is one of the predictions that will be tested in this experiment.

The research aims of this study are twofold. The first priority is to explicitly assess the contributions of remembering and knowing to the recognition of faces and voices. Methodologically, this study diverts from the previous work done by Mäntylä (1997) and Brandt et al (2003) in terms of the test materials used. Whilst these studies have used newly learnt faces as test materials, the current experiment will focus on the recognition of pre-experimentally familiar people and will therefore use famous faces and famous voices. The experiments reported in this chapter will use the selection of famous and non-famous faces and voices that have been used throughout this thesis. This will mean that in the standard yes/no recognition phase of the task, the participant has to decide whether the face or voice presented to them belongs to somebody they have encountered previously in the media (famous) or not (non-famous). As with most standard R/K studies, if a participant answers 'yes' in this initial recognition phase, they will also be required to make a remember, know or guess judgment. The addition of the guess response was chosen in light of Gardiner et al's (2002) arguments that have been discussed in earlier sections.

The second and perhaps more central aim of the experiment is to investigate the role of subjective awareness when overall recognition between faces and voices is matched. One obvious problem with recognizing famous voices relative to famous faces is likely to be exacerbated by the fact that we probably encounter a famous person's face more often in the media than we hear their voice. In order to make the hit and false alarm rate similar in the two modalities, face recognition will be reduced by presenting the faces in a blurred format. This format was chosen on the basis of the results from a pilot study which compared the effectiveness of different degradation techniques (see Chapter 5) on face recognition. Furthermore, using blurred faces adds ecological value to this study. A blurred face is a naturally occurring form of degradation which has been likened to viewing a face from a distance (Lander et al, 1999). Therefore, the three experimental conditions in this experiment will be the standard faces, voices and blurred faces.

The findings from Hanley et al's (1998) study and the experiments in Chapter 3 showed that there were more instances when familiar faces were associated with contextual retrieval relative to voices that were found familiar. In the context of the current experiment, this pattern may appear in the form of more remember responses to faces than voices. This is because remember responses, by definition, relate to the amount of contextual information attributed to the studied item.

There may be changes in R/K responses from standard faces to blurred faces, however the central and most important comparison in this study is between faces and voices. As already discussed, the main aim of designing the blurred condition is to match overall recognition with the voice condition. Whilst it has already been predicted that know responses may be strongly associated with the voice condition, will this also be the case in the blurred condition? If dissociations between remembering and knowing become apparent when overall recognition is matched (blurred condition vs voice condition) this would add to the growing body of research that has already demonstrated such dissociations both across different conditions (Mäntylä, 1997) and groups (Perfect et al, 1995). More importantly, if know responses are more prevalent in the voice condition than in the blurred face condition, then the argument outlined in Chapter 3, that contextual information such as occupations, can be more readily associated with faces than voices would be strongly supported.

Whilst this study will adopt the standard procedure used in most R/K studies, this experiment will closely follow the descriptive approach adopted by Gardiner et al (1998). In particular, after participants have completed their R/K judgments, they will be asked to justify their initial recognition decision. Participants will be probed to provide as much information as possible in order to ascertain the degree of occupational information they can provide to an item that they had recognized. It could be argued that such a technique may in fact spuriously increase the amount of detail provided, especially since the recognized item will be re-presented to the participant for this second phase of the experiment. However, the main aim of this phase in the experiment is to examine whether participants are using remember, know and guess responses in a valid way. If this is the case, then there should be more occupational information given to remember responses than know responses. Occupational information should also be more strongly associated with know responses than guess responses.

In addition to examining the role of subjective awareness in person recognition, this experiment will also assess memory sensitivity, as measured by the signal detection measures of A'. Previous studies that have calculated A' measures have shown that remember and know responses show memory for studied and non-studied items, whereas guess responses were equally likely to show memory for studied and non-studied items (Gardiner et al 2002). In the context of this experiment, remember and know responses should produce A' prime values which are significantly above chance, as both types of responses should show memory for famous items over non-famous items. Guess responses, on the other hand, should produce A' measures which do not differ significantly from the 0.50 value. This in turn would show no memory for famous items over non-famous items.

As well as looking at the A' measures for remember, know and guess responses separately, A' measures will also be calculated to examine whether knowing provides an additional source of memory. According to Gardiner et al (2002), when know responses are separated from guess responses, then knowing does provide an additional source of memory. This is reflected by A' values being significantly greater when calculated from remember and know responses, than when calculated from remember responses alone. If A' values obtained in this study are consistent with those found by Gardiner et al (2002) then this would demonstrate that know responses do provide an additional source of memory.

The main aim of this study is to contribute to the new body of research that has attempted to integrate the R/K procedure to person recognition by looking at the role of subjective awareness into the overlooked area of familiar face and voice recognition.

## Method.

### *Experiment 4a.*

**Participants.**

A total of 54 Essex University students volunteered to take part in the experiment, many in partial fulfillment of course credit. Those that were not eligible for course credit were paid £5.00 for their participation. The sample consisted of 30 males and 24 females with a mean age of 23.4 years. All participants had normal to corrected vision and normal hearing. All were Native English speakers who had spent the past 15 years living in the UK. Each participant was randomly allocated to one of three modality groups. In total, there were 18 participants per group.

**Apparatus.**

Presentation of the stimuli was by means of an Apple Mac G4 computer running QuickTime Media software. A 17 inch flat screen LCD colour monitor was used to present the visual stimuli (standard faces and blurred faces). The QuickTime Media program was also used to present the voices through a pair of Sennheiser headphones. The computer's keyboard was used by the experimenter to control the presentation rate of the stimuli.

**Stimuli.**

The stimuli included the same selection of 50 famous items and 50 non-famous items that have been used throughout this thesis. The experimental phase consisted of 48 famous items and 48 non-famous items, the remainder of the stimuli were used for the practice trials. The experimental trials were arranged into two blocks of 48 trials per block (24 famous and 24 non-famous). Half the participants completed Block 1, followed by Block 2, whilst the remaining half completed Block 2, followed by Block 1. A full list of the celebrities used in the experiment can be found in Appendix 3.

*Standard Face condition:* All faces (famous and non-famous) were edited through the multi-media software program Final Cut Pro to remove contextual background and clothing cues (see Chapter 2 for more details). The presentation of the faces (without audio) was via the software program QuickTime Player.

*Blurred Face condition:* The same selection of stimuli that were used in the Standard Face condition were re-edited through the application of a Gaussian blur filter in Final Cut Pro. All clips were rendered with a blur value of 13. This value is arbitrary and was selected by the experimenter as it produced satisfactory levels of degradation. Furthermore, the pilot study conducted in Chapter 5, showed that this level of blur was sufficient to reduce overall levels of recognition to a comparable rate to voices. The blurred face condition was then exported, without sound, as a QuickTime file.

*Voice condition:* Both famous and non-famous voices were formatted as an audio QuickTime file for presentation. The stimuli used were carefully selected to control for semantic content as well as perceptual cues such as background noise. A detailed outline of the voice selection procedure can be found in Chapter 2.

The duration of each clip was approximately 7 seconds. This level was the same across all three modality conditions. On the basis of the analysis conducted in Chapter 3 (see pages 94 –96), all items (famous and non-famous) that were used were British.

### Design.

The experiment used a mixed factorial design with one between subjects factor (modality: standard faces, blurred faces, voices) and one within subjects factor (subjective awareness: remember, know and guess). As addressed in Chapter 3 with familiarity judgments and by Brandt et al (2003) and Mäntylä (1997) with R / K and G (guess) judgments, subjective awareness is not methodologically an independent variable. However, statistical comparisons between these responses are integral to the theoretical aims of the study, therefore subjective awareness was treated as an additional factor.

One of the dependent measures was the proportion of correct (hits) and incorrect (false alarms) responses made to remember, know and guess responses as a function of modality. The second dependent measure was the proportion of correct occupational descriptions given in relation to remember, know and guess responses as a function of modality. This latter dependent variable was used to assess the validity of the Remember, Know and Guess responses.

### Procedure.

There were two testing phases in the experiment; the initial recognition phase and the occupational description phase. Details of the initial recognition phase are outlined below.

#### *Phase 1: YES/NO recognition.*

Participants entered a quiet testing room and were seated approximately 70 cm from the computer monitor. In order to prevent the resolution of the blurred faces, participants were specifically instructed in this condition to refrain from moving in their seat.

Participants were informed that they would be presented with a sequence of standard faces / blurred faces / voices, depending on the condition to which they were assigned to. They were informed that some of the faces/voices presented to them belonged to people who were well known in the British media, whereas the other faces/voices belonged to non-famous individuals. For the initial recognition phase, participants had to decide, for each item presented to them whether they recognized it or not by saying 'yes' or 'no', respectively. If they had answered 'yes', they were instructed to provide a remember, know or guess response.

The instructions for these responses were modeled on those used by Gardiner et al (1998) and are provided in Appendix 4. Before the initial testing phase began participants were asked if they fully understood the distinctions between the three types of responses. All participants were required to give a verbal response for each item, along with a pointing response to one of the 'remember', 'know' and 'guess' card laminates that were laid out in front of them. This pointing response was included to minimize any confusions between subjects responding 'no' and 'know'.

After the presentation of each item, the clip was paused for approximately 20 seconds to enable the participants to make their response. Presentation of the 96 experimental trials were in a fixed random order. The remaining four trials were used as practice trials. Halfway through the testing phase, participants were given a short break. Each participant's response was recorded in a response booklet by the experimenter.

### *Phase 2: Occupational descriptions.*

The aim of this phase of the experiment was to examine whether participants were using remember, know and guess responses in a valid way.

After the initial testing phase was completed, the experimenter randomly selected a proportion of 'yes' responses. These responses comprised of approximately equal selection of remember, know and guess responses. However, in some cases, such as in the Standard Face condition, equal selection proved difficult, because of the disproportionate number of remember, know and guess responses.

The selection was made blindly with respect to whether the response was correct or incorrect. For this phase of the experiment, the selected items were re-presented to the participant. Following Gardiner et al's (1998) recommendation, participants were not pressurized to justify their remember, know or guess responses. This was done so that the opportunity for confabulation is reduced. Instead of justifying their subjective awareness decisions, the participants were asked what it was about the item that led them to recognize it as an item they had encountered previously. Emphasis was therefore on the initial recognition decision rather than on their decision in subjective awareness.

Participants were not explicitly asked to provide a name or occupation to the face/voice, but were probed to describe in as much detail as possible what it was about the item that made them believe they and encountered it before. Specifically, participants were asked *where* and *when* they thought they had encountered the face/voice before. Probing the participants in this way ensures that the form of questioning is consistent with the remember and know definitions.

After both testing sessions were completed participants were de-briefed on the aims of the experiment. As part of the de-briefing process, participants were given the list of names of the famous people they either heard or saw in the experiment. They were asked to place a tick next to each name if they knew who the person was outside of the context of the experiment. This was done to ensure that all participants were familiar with the famous items that were presented to them.

The two testing phases and the de-briefing phase took approximately 1 hour to complete. Once the de-briefing phase was finished, participants were thanked for their time and paid for their participation.

## Results.

There are a number of different dependent measures of interest in the current study and the results section is designed around the following structure to address these issues systematically. The results section begins with an analysis of the overall recognition performance across the three modality conditions. The specific aim of this analysis is to confirm that blurred faces produced similar hit and false alarm values to voices.

The second part of the results section decomposes the overall recognition rate into the relevant subjective awareness categories (remember, know and guess). The aim of this section of the analyses is to examine whether different modalities produce different effects on subjective awareness.

The third section aims to objectively assess the validity of R, K and G responses by analyzing the corresponding A' values for each category. The comparison of A' values for R responses versus A' values for R and K responses combined was used to assess some of the assumptions of the unitary strength model (Donaldson, 1996).

The final section of the analysis attempts to subjectively assess the validity of R, K and G responses. The descriptive approach adopted by Gardiner et al (1998) showed how the level of elaborate contextual detail decreased as responses changed from R to K to G responses. The analysis of occupational descriptions across the three subjective awareness levels aimed to obtain similar findings.

### *1./ Overall recognition.*

For each participant, the hit and false alarm rates were calculated by dividing the number of correct responses (hits) by 48 and the number of incorrect responses by 48 (false alarms) and then averaged across participants to produce the means and standard deviations in table 6.1.a

Table 6.1.a: Mean overall hits, false alarm rates and their corresponding A' (sensitivity) and B D (bias) measures for the standard face, blurred face and voice conditions. Standard deviations in parentheses.

| **Condition** | **Hits** | **False Alarms** | **A'** | **B"D** |
|---|---|---|---|---|
| Standard Faces | 0.90 (0.08) | 0.13 (0.07) | 0.94 (0.04) | -0.19 (0.55) |
| Blurred Faces | 0.70 (0.10) | 0.21 (0.12) | 0.83 (0.06) | 0.21 (0.44) |
| Voices | 0.69 (0.12) | 0.27 (0.10) | 0.80 (0.07) | 0.10 (0.40) |

A series of independent samples t-test was carried out on the hits and false alarms values to establish which condition produced the best performance and importantly to assess whether the performance between the blurred face condition and voice condition was matched.

The results showed that there were no significant differences in terms of hits ($p=0.811$) and false alarms ($p=0.644$) in the blurred and voice conditions. The hit rates produced in the standard face condition were significantly higher than both the blurred face, $t(34) = 6.48$, $p < 0.05$ and the voice condition $t(34) = 6.24$, $p < 0.05$, respectively. False alarm rates were significantly lower in the standard face condition compared to the blurred condition, $t(34) = 2.74$, $p < 0.05$ and the voice condition $t(34) = 4.26$, $p < 0.05$, respectively. The same conclusions were drawn from the analyses of the A' data. This showed that the A' value for the standard face condition was significantly higher than blurred faces and voices ($p < 0.05$), which in turn did not differ from each other ($p > 0.05$). There were no differences in bias across the three modalities ($p > 0.05$).

This suggests that participants in the standard face condition were better at discriminating famous items from non-famous items compared to the participants in the blurred and the voice conditions – who in turn were more likely to incorrectly categorize a non-famous item as a famous item.

More central to the research objectives of this study are the results comparing the performance between the blurred face and voice conditions. The non-significant differences in hit and false alarm rates for these two conditions illustrates that presenting the faces in a blurred format was successful in making the overall recognition of faces the same as that in the voice condition.

### *2.1 Subjective awareness as a function of modality.*

The second analysis examines the relation between modality and subjective awareness (see table 6.1.b).

Table 6.1.b: Mean proportion of correct (hits) and incorrect (false alarms) remember, know and guess responses for the standard face, blurred face and voice conditions. Standard deviations in parentheses.

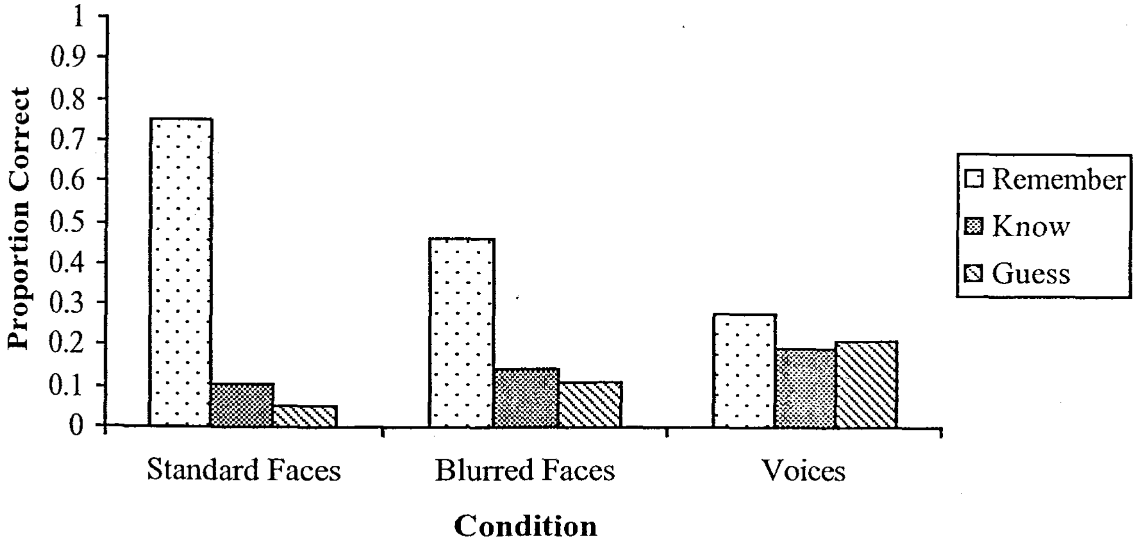
| | **Hits** | | | **False Alarms** | | |
|---|---|---|---|---|---|---|
| **Condition** | Remember | Know | Guess | Remember | Know | Guess |
| Standard Faces | 0.75<br>(0.15) | 0.10<br>(0.05) | 0.05<br>(0.07) | 0.00<br>(0.01) | 0.04<br>(0.03) | 0.08<br>(0.05) |
| Blurred Faces | 0.46<br>(0.15) | 0.14<br>(0.07) | 0.10<br>(0.08) | 0.02<br>(0.02) | 0.07<br>(0.07) | 0.12<br>(0.09) |
| Voices | 0.27<br>(0.16) | 0.20<br>(0.07) | 0.22<br>(0.09) | 0.01<br>(0.01) | 0.08<br>(0.05) | 0.18<br>(0.07) |

This analysis decomposed the overall hit values for each participant into remember, know and guess responses. The hit values were then submitted to a 3x3 mixed ANOVA. There was one between-subjects factor (modality: standard faces, blurred faces and voices) and one within subjects factor (subjective awareness: remember, know and guess). This analysis revealed a main effect of modality, F (2,51) = 23.49, $p < 0.05$ and a main effect of subjective awareness F (2, 102) = 153.64, $p < 0.05$. Furthermore, the modality by subjective awareness interaction was also significant F (4, 102) = 36.17, $p < 0.05$, (see table 6.1.b for treatment means).

The analyses most relevant to the predictions are the comparisons between remember and know responses across the three modalities. Starting with the blurred face and voice conditions, planned comparisons revealed significantly more R responses to blurred faces than voices ($p < 0.05$). Furthermore, there were significantly more K responses to voices than blurred faces ($p < 0.05$) Planned comparisons also revealed significantly more remember responses made to standard faces compared to blurred faces, ($p < 0.05$) and voices ($p < 0.05$). The analysis of know responses showed that blurred faces were significantly more likely to be associated with know responses than standard faces, ($p < 0.05$) and there were also more know responses made to voices than standard faces, ($p < 0.05$). This pattern of results is presented in figure 6.1.a, which summarizes two important aspects of the data.

Firstly, it shows how changing the presentation format of faces has different effects on subjective awareness. In this case, presenting the same faces in a blurred format decreases remembering whilst increasing knowing. However, the critical findings are observed in the difference between the blurred face and voice conditions. As can be seen in table 6.1.a, the hit and false alarm rates in the blurred and voice conditions are virtually identical, yet the two modalities produce selective effects on subjective awareness. As figure 6.1.a illustrates, there are more remember responses made to blurred faces than to voices, whereas more know responses are associated with voices than blurred faces.

Figure 6.1.a: Proportion of correct Remember, Know and Guess responses for each modality condition.



*3.1 Signal Detection Measures.*

The first set of signal detection measures represent the A' values for remember, know and guess responses calculated separately. According to Gardiner et al (2002), if participants are categorizing their levels of subjective awareness in the correct way, then A' values for R and K responses should be significantly above chance. Guess responses, on the other hand, should not statistically differ from chance.

A' values range from 0 (no discrimination) to 1.0 (perfect discrimination). Chance performance is represented by the A' value of 0.50. Measures of bias was also calculated for each of the measures by using B"D, where –1.00 represents lenient responding and +1.00 represents conservative responding.

Table 6.1.c: Mean sensitivity (A’ Prime) and criterion measures (B”D Prime) calculated separately for remember, know and guess responses. Standard deviations in parentheses.

| | **A’** | | | **B D** | | |
|---|---|---|---|---|---|---|
| **Condition** | Remember | Know | Guess | Remember | Know | Guess |
| Standard Faces | * 0.94<br>(0.04) | * 0.61<br>(0.20) | 0.43<br>(0.17) | 0.96<br>(0.11) | 0.94<br>(0.21) | 0.99<br>(0.02) |
| Blurred Faces | *0.85<br>(0.04) | * 0.64<br>(0.16) | 0.48<br>(0.13) | 0.95<br>(0.07) | 0.97<br>(0.04) | 0.95<br>(0.08) |
| Voices | *0.81<br>(0.05) | * 0.67<br>(0.09) | 0.55<br>(0.13) | 0.99<br>(0.03) | 0.95<br>(0.05) | 0.87<br>(0.11) |

* = significantly greater than chance ($p < 0.05$).

The series of one sample t-tests, which used 0.50 as the test value, produced results which were consistent with Gardiner et al’s arguments. Specifically, Remember and Know responses produced A’ values that were significantly above chance in all of the modality conditions. Furthermore, each of the A’ values for Guess responses did not differ significantly from chance.

The second set of A’ values aims to address one of the assumptions of the unitary strength model (Donaldson, 1996). For the unitary strength model to be upheld (Donaldson, 1996), A’ measures calculated from remember only hits and false alarms should not differ significantly from the A’ value calculated from remember plus know hits and false alarms. This is based on the assumption that the inclusion of know responses do not provide an additional source of memory. The following table includes A’ and B”D values calculated from remember only responses and remember plus know responses. The assumptions of the unitary strength model are tested by carrying out a series of paired samples t-test on the relevant A' prime values.

Table 6.1.d: Mean sensitivity (A') and criterion measures (B"D) for remember only and remember plus know data for the three modality conditions. Standard deviations in parentheses.

| | **Remember Only** | | **Remember + Know** | |
|---|---|---|---|---|
| **Condition** | A' Prime | B D Prime | A' Prime | B D Prime |
| Standard Faces | 0.94 (0.04) | 0.96 (0.11) | 0.95 (0.03) | 0.43 (0.46) |
| Blurred Faces | 0.85 (0.04) | 0.95 (0.07) | 0.86 (0.04) | 0.66 (0.36) |
| Voices | 0.81 (0.04) | 0.99 (0.03) | 0.80 (0.06) | 0.72 (0.28) |

For the standard face condition, A' measures for remember plus know responses were significantly greater than the A' prime measure for remember responses alone, t (17) = 2.80, p = 0.006 (one tailed).

For the blurred face condition, the A' prime measure for remember plus know responses did not differ significantly from the A' prime measure for remember responses alone (p = 0.286, one tailed).

For the voice condition, the remember plus know A' prime measure did not differ significantly from the A' prime measure for remember only responses (p = 0.395, one tailed).

From these analyses, the standard face condition was the only condition in which knowing provided an additional source of memory. This particular outcome does not support the assumptions of the unitary strength model by demonstrating that know responses can be used to improve recognition performance.

### *4.1 Descriptions of remembering, knowing and guessing.*

The following analysis attempts to examine the type of descriptions participants give when they claim to have recognized the face/voice. This type of analysis follows on from Gardiner et al's (1998) descriptive approach to the R/K/G paradigm. The main area of interest in this set of analyses is whether the retrieval of occupational details differs across the three levels of subjective awareness.

*Scoring.*

Following Gardiner et al's (1998) descriptive approach to the R/K/G paradigm, some examples of the kinds of descriptions given for each state of awareness is provided in Appendix 5.

Analysis of the descriptions focused more on the detail related to occupational information rather than the ability to name the person. Occupational descriptions were favoured over name recall primarily because it is generally easier to retrieve an occupational description about a person than it is to name them (Young, McWeeny, Ellis and Hay, 1986), which in turn would allow for a greater selection of responses.

A description is deemed correct if it contained some form of identity specific information. For example, the description for Joanna Lumley, 'used to be in Avengers, now seen in Absolutely Fabulous' is scored as a correct description whereas 'on television' or 'actress' is not.

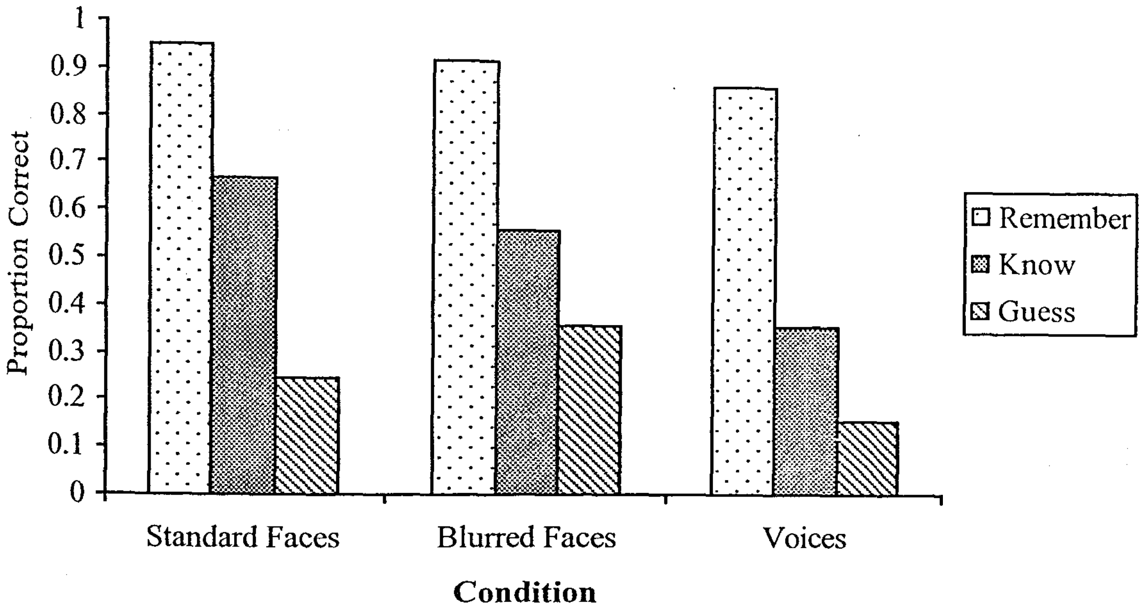
Each participant's occupational score was calculated by dividing the number of correct descriptions by the number of selections made at each level of subjective awareness. For example, if a participant is asked to provide descriptions for 8 remember responses, 8 know responses and 8 guess responses and subsequently provides 7, 4 and 2 correct descriptions respectively, then that participant's occupational score would be 0.88 for remember, 0.50 for know and 0.25 for guess responses.

Table 6.1.e: Mean proportion of occupational descriptions given to remember, know and guess responses across the three modality conditions. Standard deviations in parentheses.

| | **Occupational Descriptions** | | |
|---|---|---|---|
| **Condition** | Remember | Know | Guess |
| Standard Faces | 0.95 (0.05) | 0.66 (0.26) | 0.25 (0.15) |
| Blurred Faces | 0.92 (0.08) | 0.57 (0.35) | 0.36 (0.26) |
| Voices | 0.87 (0.11) | 0.36 (0.28) | 0.16 (0.15) |

These occupational scores were calculated for each participant in each of the three conditions and submitted to a 3 (modality) x 3 (subjective awareness) mixed ANOVA (see table 6.1.e for treatment means). This revealed a main effect of modality $F(2, 51) = 7.03$, $p < 0.05$ and a main effect of subjective awareness $F(2, 102) = 178.36$, $p < 0.05$. The modality by subjective awareness interaction was also significant $F(4, 102) = 3.02$, $p < 0.05$. The nature of this interaction was investigated further by a series of planned comparisons t-tests which aimed to assess the differences in occupational retrieval across the three levels of subjective awareness. These analyses showed that for each of the three modality conditions there were significantly more occupations given to remember responses, than know responses and significantly more occupations given to know responses than guess responses ($p < 0.05$). This pattern of performance is presented in figure 6.1.b.

Figure 6.1.b: Proportion of occupations retrieved for Remember, Know and Guess responses as a function of modality.



## Discussion.

The aim of this study was to apply the R/K/G procedure to the area of person recognition. Previous work that has attempted to integrate this paradigm in person recognition memory has used newly learnt faces (Mäntylä, 1997 and Brandt et al, 2003). This study extends this research to familiar faces and voices.

The series of analyses conducted on the R/K/G responses showed that participants understood the instructions and applied each response in the correct way. Using the objective value of A', analysis showed that only Remember and Know responses produced recognition performance that was significantly above chance. These particular observations are consistent with earlier findings obtained by Gardiner et al (1998). In addition to the objective measures taken, participants' occupational descriptions are also consistent with R/K/G responses. Specifically (see Appendix 5), there were significantly more occupations retrieved to remember responses than know responses, which in turn were associated with more occupational information than guess responses. This pattern of results validates Gardiner et al's (1998) descriptive approach to assessing R/K/G responses and in doing so, elegantly illustrates how the three types of responses vary in contextual information (see figure 6.1.b).

Throughout this study, there was considerable theoretical emphasis placed on producing a matched control for the face recognition. This was imposed in order to assess the relative contributions of Remembering and Knowing to face and voice recognition when overall recognition between the two modalities is matched. This was achieved experimentally by presenting the faces in a blurred format. As table 6.1.a illustrates, the blurred condition produced a hit rate that was slightly higher than voices, however, the statistical analyses on both the hit and false alarm rates showed that blurred faces and voices did not differ significantly from each other.

The critical aspect of the analyses then moved onto to assess the R and K responses across the three conditions. It is worth to point out at this point, that guess responses were not examined across the three modality conditions. This was because there were no *a priori* predictions to make with regards to this response type. Moreover, guess responses were used in this study purely for theoretical purposes as outlined by Gardiner et al (2002). Specifically, guess responses were incorporated into the R/K paradigm to prevent participants from using K responses if they were unsure.

Returning to the R/K responses, the critical area of interest is between the blurred faces and voice condition. It is more interesting to compare performance between the two modalities when overall recognition is matched. However, before these conditions are compared there are few points that need to be made with regards to the pattern of performance with the standard face condition. The standard face condition was associated with more remember responses than the blurred face and voice condition. On the other hand, know responses were higher in the blurred face condition than in the standard face condition. Similarly, there were more know responses to voices than blurred faces. Returning to the main data of interest, blurred faces and voices did not differ from each other at a quantitative level, however how did the two modalities compare at a qualitative level?

As illustrated in figure 6.1.a, blurred faces and voices produce selective effects on remembering and knowing. This is in the form of more remember responses made to blurred faces than voices. Know responses were more likely to be associated with voices than blurred faces. This suggests that faces and voices are recognized in qualitatively different ways. Although this is a theoretically interesting finding, caution is warranted because the trace-strength account provides a potential alternative explanation.

A proponent of the trace strength account (Donaldson, 1996) would argue that qualitative differences between blurred faces and voices would reflect nothing more than differences in trace strength or subjective confidence. A way this has been tackled in the R/K literature is by re-running the experiment and asking participants to follow up their yes/no recognition decisions with a confidence rating instead of R/K/G responses. This would be an important experiment to run, as preliminary inspection of figure 6.1.a seems to suggest a trace-strength account for the data.

This can be seen in the comparison of guess responses between blurred faces and voices. As figure 6.1.a shows, there is a higher proportion of guess responses made to voices than blurred faces. This observation is also confirmed statistically, with a *Scheffé* test indicating that the difference between the two modalities is significant ($p \leq 0.05$). If the qualitative differences observed between blurred faces and voices can be summarized simply in terms of differences in subjective confidence, then participants in the blurred face condition should be more confident than participants in the voice condition. This particular prediction was investigated in experiment 4b below.

## Method.

### *Experiment 4b.*

**Participants.**

Thirty Psychology Open Day students and staff who were visiting the Department of Psychology participated in the experiment. The mean age of the participants was 21.7 years. The sample consisted of 20 males and 10 females. There were 15 participants allocated to the voice condition and 15 participants allocated to the blurred face condition. All participants had normal hearing and normal to corrected vision. All of the participants were Native English speakers who had resided in the U.K for the last 15 years. Participants were given a book voucher for their participation.

**Apparatus.**

Same as experiment 4a with the addition of a Panasonic CD player for presentation of voice stimuli.

**Stimuli.**

Same as experiment 4a.

**Design.**

The experiment used a two factor mixed design. Modality was a between subjects factor, which included blurred faces and voices. Confidence was a within subjects factor, which included the confidence levels very sure, less sure and guess. The main dependent measure of interest in this study is the proportion of correct responses (hits) at each level of confidence across the two modalities.

**Procedure.**

Apart from the test instructions, the stimuli and the presentation order of the stimuli were identical to those of experiment 4a. The procedural elements of this experiment differed from experiment 4a on the following grounds. In terms of the instructions, participants were provided with exactly the same instructions for the yes/no recognition task (phase 1) as used in experiment 4a. However, instead of following up their 'yes' response with an R/K/G response, the participants were required to make confidence judgments for each recognized item. For their confidence judgments participants were provided with the following scale: Very sure, Less sure and Guess.

Due to time and resource constraints, participants were not tested individually but in groups of 15 participants per group. Each participant was seated in front of an Apple Macintosh computer and was shown how to use the QuickTime Player file and how to use the spacebar to move on to the next clip in the sequence. Participants in the blurred face condition were instructed to remain as still as possible in their seats. Participants in the voice condition were presented with each voice item through speakers connected to a Panasonic CD player. Presentation rate was controlled by the experimenter. Unlike experiment 6b, participants were required to note down their responses in the answer booklet provided to them.

As this experiment was primarily concerned with the distribution of confidence judgments, participants were not subsequently tested with phase 2 of the experiment.

## Results.

The first set of analyses looks at the overall levels of performance across the three modality conditions. The hits and false alarms were calculated in the same way as in the previous experiment. Table 6.2.a shows the proportion of hits and false alarms for the three modality conditions.

Table 6.2a: Mean overall hits, false alarm rates and their corresponding A' (sensitivity) and B D (bias) measures for the blurred face and voice conditions. Standard deviations in parentheses.

| **Condition** | **Hits** | **False Alarms** | **A'** | **B"D** |
|---|---|---|---|---|
| Blurred Faces | 0.61 (0.20) | 0.20 (0.14) | 0.81 (0.05) | 0.42 (0.44) |
| Voices | 0.55 (0.08) | 0.32 (0.09) | 0.69 (0.06) | 0.28 (0.30) |

Extending the findings from experiment 4a, the hit rates between the blurred face and voice condition did not differ significantly ($p > 0.05$). However, contrary to expectation, the false alarm rate in the voice condition was found to be statistically higher than in the blurred face condition, $t(28) = 2.81$, $p < 0.05$. Both blurred faces and voice conditions produced hit rates that are somewhat lower than those observed in the R/K/G study (see table 6.1.a). The A' value for the blurred face condition was also higher than in the voice condition $t(28) = 5.74$, $p < 0.05$. The measures of bias (B"D) did not significantly differ between the two conditions ($p > 0.05$).

The overall recognition data were then decomposed into the proportion of hits and false alarms for each confidence category across the two modality conditions. These values are displayed table 6.2.b.

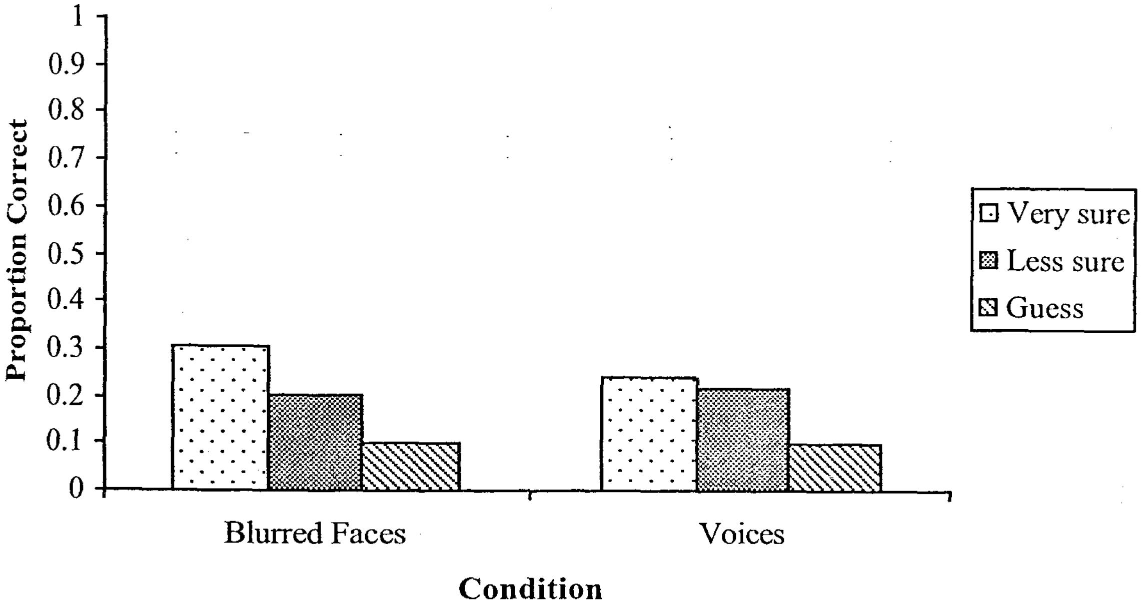
Table 6.2.b: Mean proportion of correct and incorrect very sure, less sure and guess responses for the blurred face and voice conditions. Standard deviations in parentheses.

| | **Hits** | | | **False Alarms** | | |
|---|---|---|---|---|---|---|
| **Condition** | Very sure | Less sure | Guess | Very sure | Less sure | Guess |
| Blurred Faces | 0.30<br>(0.16) | 0.20<br>(0.08) | 0.10<br>(0.08) | 0.02<br>(0.03) | 0.07<br>(0.07) | 0.10<br>(0.09) |
| Voices | 0.24<br>(0.09) | 0.22<br>(0.10) | 0.10<br>(0.09) | 0.04<br>(0.05) | 0.17<br>(0.12) | 0.10<br>(0.08) |

The aim of this experiment was to test the trace-strength hypothesis (Donaldson, 1996). Under this account, the high association of remember hit responses observed for blurred faces over voices in the previous experiment could be due to greater confidence in the blurred face condition than in the voice condition. If the dissociation between remember and know responses observed for blurred faces and voices in the previous experiment can be explained away simply in terms of greater confidence for blurred faces, then this should be reflected by a significant modality by confidence interaction.

The 2x3 ANOVA failed to produce a main effect of modality, $F(1, 28) = 0.58$, $p > 0.05$. There was a main effect of confidence, $F(2, 56) = 17.81$, $p < 0.05$. However, the modality by confidence interaction was not significant, $F(2, 56) = 1.11$, $p > 0.05$. This pattern of performance is displayed in figure 6.2.a.

Figure 6.2.a: Proportion correct for very sure, less sure and guess responses for the blurred face and voice conditions.



### Discussion.

Although not originally part of the analyses of experiment 4a, guess responses were found to be significantly higher in response to voices than blurred faces. This particular finding would suggest that the reason for the higher levels of remember responses to blurred faces than voices is simply due to participants showing higher levels of subjective confidence for blurred faces compared to voices.

This trace-strength (Donaldson, 1996) account was explicitly tested in this experiment by asking participants to follow up their 'yes' responses with a confidence judgment. Contrary to the trace-strength account, the pattern of results shown in figure 6.2.a suggests that recognition for blurred faces is not associated with greater confidence than voices. This is consistent with previous research which has shown that remember responses are not simply a product of trace strength (Gardiner and Java, 1990, Mäntylä, 1997).

The main area of concern in the current study was the overall low level of performance in the voice condition. Unlike experiment 4a, blurred faces and voices were not matched for false alarm values and this can be clearly seen in the significant difference in the A' values between these two conditions. Therefore experiment 4a was repeated with the aim of matching overall performance between blurred faces and voices. Furthermore, participants were only tested with blurred faces and voices with the aim of replicating the observed dissociation between remember and know responses across the two modalities.

Although experiment 4a suggested that the recognition of blurred faces was somewhat easier than voices, this experiment discounted this hypothesis by assessing participants' confidence judgments across the two modalities. A more stringent way of assessing the dissociation between R and K responses across blurred faces and voices is to reduce the level of recognition in the blurred face condition even further. This will be done in experiment 4c by increasing the blur level for faces. This particular manipulation would serve as an alternative way of assessing the trace-strength account. Specifically, if blurred faces and voices differed in terms of subjective confidence, then voice recognition should (again) be associated with a higher proportion of guess responses than blurred faces. If the level of guessing is the same across the two modalities, but the dissociation between remember and know responses between faces and voices remains, then this would provide unequivocal support for the argument that blurred faces and voices entail qualitatively different memory processes.

## Method.

### *Experiment 4c.*

### Participants.

A total of 36 Essex University students volunteered to take part in the experiment, many in partial fulfillment for course credit. Those that were not eligible for course credit were paid £5.00 for their participation. The sample consisted of 12 males and 24 females with a mean age of 21.8 years. All participants had normal or corrected to normal vision and normal hearing. All were Native English speakers who had spent the past 15 years living in the UK. Each participant was randomly allocated to one of two modality groups. In total, there were 18 participants per group.

### Apparatus.

Same as Experiment 4a.

### Stimuli.

*Voice condition*: the number and selection of stimuli was the same as experiment 4a.

*Blurred condition:* the number and selection of stimuli was the same as experiment 4a. However the degree of the Gaussian blur filter was increased from 13 in experiment 4a to 26.25 for this experiment. Again, this value is arbitrary and was selected by the experimenter as it produced a noticeable deterioration in the quality of the image whilst still retaining some face-like properties. An increase above this 26.25 value was deemed too coarse for recognition purposes. Specifically, blur values above this level tended to obliterate even the outer headline.

## Design.

The experiment used a mixed factorial design with one between subjects factor (modality: blurred faces and voices) and one within subjects factor (subjective awareness: remember, know and guess). The dependent measures were the same as experiment 4a.

## Procedure.

Same as Experiment 4a.

## Results.

The results section for this experiment follows the same structure as the results section in experiment 4a.

### *1./ Overall recognition.*

Table 6.3.a: Mean overall hits, false alarm rates and their corresponding A' (sensitivity) and B D (bias) measures for the blurred face and voice conditions. Standard deviations in parentheses.

| Condition | Hits | False Alarms | A' | B"D |
|---|---|---|---|---|
| Blurred Faces | 0.63 (0.13) | 0.24 (0.14) | 0.78 (0.08) | 0.27 (0.42) |
| Voices | 0.67 (0.14) | 0.28 (0.11) | 0.79 (0.06) | 0.08 (0.47) |

Importantly, two independent samples t-tests showed that the hit and false alarm rates between the two conditions did not statistically differ ($p > 0.05$). Blurred faces and voice also did not differ in terms of sensitivity ($p > 0.05$) or bias ($p > 0.05$).

An additional set of analyses was conducted on the A' values for the blurred face and voice condition from this experiment with the A' values for the same conditions in experiment 4a (see table 6.1.a)

A one sample t-test using the test value of 0.80, showed that the A' value for the voice condition in this experiment did not differ significantly from the A' value for voices in experiment 4a ($p > 0.05$). On the other hand, increasing the level of blur did have a significant effect on the ability to discriminate famous from non-famous items. Experiment 4a which used a Gaussian blur filter of a value of 13 produced an A' value of 0.83. This value was compared with the A' value obtained in the current experiment. The analyses showed that increasing the level of blur to 26.25 had a significant detrimental effect on recognition performance, t (17) = 2.54, $p < 0.05$.

### *2.1 Subjective awareness as a function of modality.*

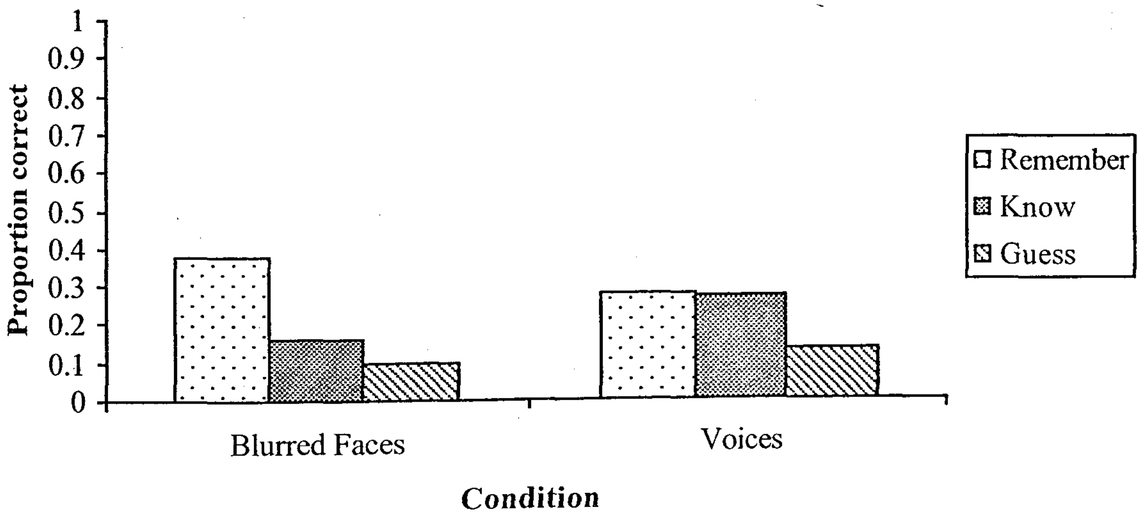
The next analyses assess the distribution of Remember, Know and Guess responses across blurred faces and voices. The overall hit and false alarm rate for each condition was decomposed into proportions of hits and false alarms for each level of subjective awareness. These values are displayed in table 6.3.b.

Table 6.3.b: Mean proportion of correct (hits) and incorrect (false alarms) remember, know and guess responses for the blurred face and voice conditions. Standard deviations in parentheses.

| | **Hits** | | | **False Alarms** | | |
|---|---|---|---|---|---|---|
| **Condition** | Remember | Know | Guess | Remember | Know | Guess |
| Blurred Faces | 0.38<br>(0.17) | 0.16<br>(0.10) | 0.10<br>(0.08) | 0.04<br>(0.04) | 0.12<br>(0.09) | 0.08<br>(0.09) |
| Voices | 0.28<br>(0.14) | 0.27<br>(0.09) | 0.13<br>(0.11) | 0.01<br>(0.02) | 0.12<br>(0.08) | 0.16<br>(0.12) |

The hit values were entered into a 2 (modality) x 3 (subjective awareness) mixed ANOVA. The main effect of modality was not significant by participants, F (1, 34) = 0.74, $p > 0.05$ or by items ($p > 0.05$). However the main effect of subjective awareness was significant, F (2, 68) = 23.41, $p < 0.05$ and so was the modality by subjective awareness interaction, F (2, 68) = 5.65, $p < 0.05$. The interaction was followed up by three planned comparisons t-tests. These analyses revealed that the only significant differences between the two modalities was between the remember and know responses. Specifically, there were more remember responses made to blurred faces than voices ($p < 0.05$) whereas voices were more likely to be associated with a higher proportion of know responses than blurred faces ($p < 0.05$). The difference in guess responses between the two conditions was not significant. This pattern of performance is displayed in figure 6.3.a.

Figure 6.3.a: Proportion of correct Remember, Know and Guess responses in each condition.



*3./ Signal Detection Measures.*

The following analyses examines whether Remember, Know and Guess responses have been used by the participants in a valid way. This will be done by obtaining A' values for each of the levels of subjective awareness for each of the modalities. According to Gardiner et al (2002) the correct application of R, K and G responses will be revealed in the A' measures. Specifically, R and K responses should produce A' measures which are significantly above chance (0.50). Guess responses on the other hand should not statistically differ from chance. Table 6.3.c shows the A’ and B”D measures for each condition.

Table 6.3.c: Mean sensitivity (A’ Prime) and criterion measures (B”D Prime) calculated separately for remember, know and guess responses. Standard deviations in parentheses.

| | **A’** | | | **B D** | | |
|---|---|---|---|---|---|---|
| **Condition** | Remember | Know | Guess | Remember | Know | Guess |
| Blurred Faces | *0.81<br>(0.06) | * 0.58<br>(0.14) | 0.56<br>(0.16) | 0.93<br>(0.08) | 0.93<br>(0.08) | 0.97<br>(0.06) |
| Voices | *0.81<br>(0.03) | * 0.67<br>(0.10) | 0.49<br>(0.15) | 0.98<br>(0.02) | 0.89<br>(0.12) | 0.90<br>(0.12) |

* = $p < 0.05$

Consistent with Gardiner et al’s (2002) arguments, a series of one sample t-tests showed that Remember and Know responses produced A’ values that significantly differed from chance. Guess responses on the other hand did not differ significantly from chance. Therefore this set of analyses objectively demonstrates that the participants in this experiment were applying the three levels of subjective awareness in a valid way.

The second use of the A’ measures is to investigate whether know responses provide an additional source of memory. If this is the case, then A’ measures calculated from Remember plus Know responses should be significantly higher than

A' values from Remember responses alone. Table 6.3.d shows the mean values for these measures.

Table 6.3.d: Mean sensitivity (A') and criterion measures (B"D) for remember only and remember plus know data for the blurred face and voice conditions. Standard deviations in parentheses.

| | **Remember Only** | | **Remember + Know** | |
|---|---|---|---|---|
| **Condition** | A' Prime | B D Prime | A' Prime | B D Prime |
| Blurred Faces | 0.81 (0.06) | 0.94 (0.08) | 0.78 (0.09) | 0.60 (0.30) |
| Voices | 0.81 (0.03) | 0.98 (0.03) | 0.80 (0.08) | 0.69 (0.23) |

Two paired samples t-test failed to reveal any significant improvements in memory performance when know responses were added to recognition performance ($p > 0.05$). This pattern of results largely replicates earlier findings with blurred faces and voices from experiment 4a (see table 6.1.d).

### *4./ Descriptions of remembering, knowing and guessing.*

The final analyses attempts to extend the findings from the objective assessments of R, K and G responses (section 3 above) by analyzing the types of descriptions associated with the three levels of subjective awareness. This subjective approach has also been used by Gardiner et al (1998) as an alternative way to assess the validity of participants' responses. The general finding from Gardiner et al's research was that the level of contextual detail varied as a function of subjective awareness. Specifically, there were more contextual details associated with Remember responses than Know responses, which in turn differed from Guess responses.

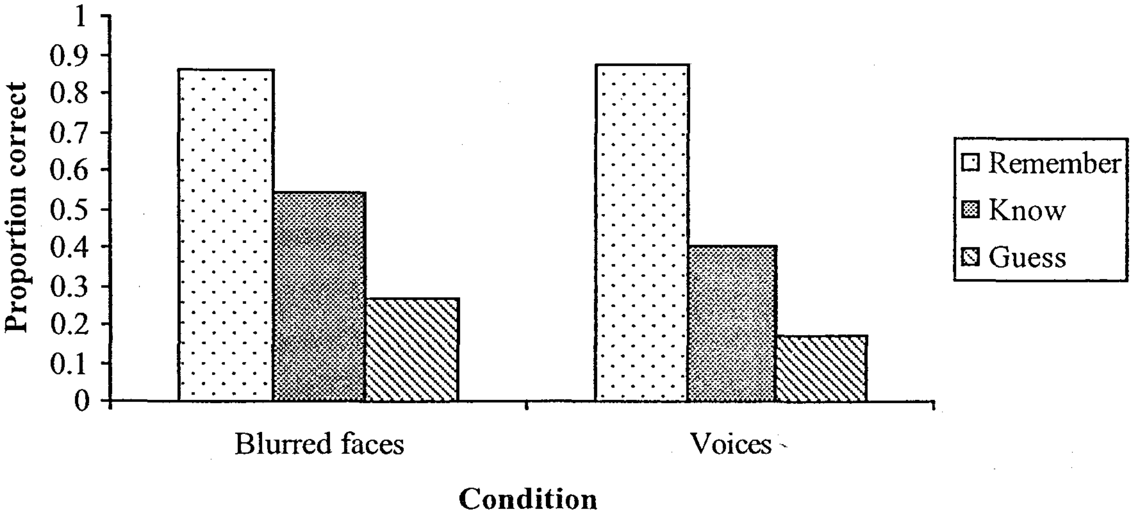
It was anticipated that a similar pattern of performance would be observed in the current experiment. Participants' descriptions in this experiment focused on the amount of occupational detail retrieved across each level of subjective awareness. The assessment and scoring of these descriptions follows the same criteria as described in section 4 in experiment 4a.

Table 6.3.e: Mean proportion of correct occupational descriptions given to remember, know and guess responses across the two modality conditions. Standard deviations in parentheses,

| | **Occupational Descriptions** | | |
|---|---|---|---|
| **Condition** | Remember | Know | Guess |
| Blurred Faces | 0.86 (0.09) | 0.54 (0.33) | 0.27 (0.21) |
| Voices | 0.88 (0.10) | 0.41 (0.34) | 0.17 (0.16) |

Each participants' occupational score was entered into a 2 (modality) by 3 (subjective awareness) mixed ANOVA. There was not a significant main effect of modality F (1, 34) = 2.41, $p > 0.05$, but the main effect for subjective awareness was significant, F (2, 68) = 74.36, $p < 0.05$. The interaction was not significant, F (2, 68) = 1.11, $p > 0.05$. Planned comparisons showed once again that there were more occupations retrieved to remember responses than know responses and more occupations retrieved to know responses than guess responses ($p < 0.05$). This pattern of performance for the occupational data is displayed in figure 6.3.b.

Figure 6.3.b: Prorportion of occupations retrieved for Remember, Know and Guess responses as a function of modality.



### Discussion.

Experiment 4c aimed to fulfil two purposes. Firstly, it aimed to validate the original findings from experiment 4a with regards to the hit and false alarm data for blurred faces and voices. Unlike experiment 4b, this was successfully achieved in the current experiment (see table 6.3.a). The second aim was to readdress the trace strength account (Donaldson, 1996). The guessing data from experiment 4a hinted at a possible difference in confidence between the blurred faces and voices. This was approached in the current experiment by making the blurred face condition harder.

The increase in the level of the Gaussian blur filter significantly reduced performance in the blurred face condition from the level of recognition obtained in experiment 4a. Critically, this level of degradation was still at a comparable level to the voice condition. In this experiment the difference in guess responses across the two modalities was not significant, thus ruling out the possibility that blurred faces are easier to recognize than voices. Importantly the dissociation between remember and know responses across the blurred face and voice condition obtained in experiment 4a, is replicated here in this experiment, even when face recognition was made harder. This suggests that face and voice recognition rely on qualitatively different memory processes to different extents.

## General Discussion.

As noted throughout this chapter, R/K instructions have always been defined to participants in person recognition terms (for example, Pefect et al, 1995). However, only a handful of studies have applied this paradigm to person recognition memory (Mäntylä, 1997, Brandt et al, 2003). Such studies have made a noteable contribution to the understanding of recognition processes involved in face recognition and how various factors such as facial distinctiveness can have selective effects on remembering (Brandt et al, 2003). The aim of this chapter was to extend this line of research to the study of familiar faces and to examine whether there are any qualitative differences between faces and voices when overall recognition is matched.

In Experiment 4a, participants were required to make remember (R), know (K) and guess (G) responses to standard faces, blurred faces and voices. Unsuprisingly, the overall recognition rate was the highest in the standard face condition, which produced near ceiling performance. As can be seen from table 6.1.a, applying a blur filter to the standard faces produced overall recognition rates that were similar to voices, both in terms of hit and false alarm rates. The development of this additional face condition served as the critical comparison for voice recognition performance. Before these two modalities are discussed in detail, it is worth highlighting the validity of the R, K and G responses.

The validity of the R, K and G responses were tested objectively in the form of A' values. These values were generated in order to test some the assumptions proposed by Gardiner et al (2002). Gardiner et al (2002) claimed that A' measures obtained for R and K responses should significantly differ from chance (0.50) because they both provide a source of memory which allows the participant to discriminate between studied (famous) and non-studied (non-famous) items. In the case of R responses, the participant can retrieve contextual information about the item, however if this fails, the participant can recognize the item beacause of strong feelings of familiarity. This level of familiarity is likely to be stronger for studied items than non-studied items. Guess responses on the other hand, may involve judgmental strategies which have little implication in the memory for the studied event. As a result, the recognition of targets and foils is roughly equal. This in turn results in an A' value that does not differ from chance. The results from table 6.1.c support these assumptions. Both R and K responses across the three modality conditions produced A' values that were significantly above chance. Guess responses, on the other hand, did not differ significantly from chance in any of the three modality conditions.

This objective assessment of R, K and G responses was complemented by a subjective assessment in the form of occupational descriptions associated with each response type. This subjective/descriptive approach to the R, K and G assessment has also been employed by Gardiner et al (1998) who showed that there were greater contextual details associated with words given an R response than a K response. Guess responses tended not to reflect correct contextual details for the studied word, but were largely based on strategic efforts coupled with low levels of confidence.

The analysis of these occupational descriptions did show that there were signifantly more occupations associated with R responses than K responses which in turn included more descriptions than guess responses. A few examples of the kinds of descriptions given by partcipants can be found in the Appendix. Therefore, this second set of analyses extends the original observations made by Gardiner et al (1998), by showing that R, K and G responses elicit different levels of contextual detail (see figure 6.1.b). Taken together, these objective and subjective assessments of

R, K and G responses demonstrate that partcipants in Experiment 4a, used these memory judgments in a consistent way.

The discussion now turns to the the distribution of R, K and G responses across the different modalities. This was assessed by breaking down the overall hit rate into the proportion of hits associated with R, K and G responses. Starting with the standard face condition, R responses were significantly higher in this condition than any of the other two conditions. This result was largely expected given the findings from previous resarch which have shown that faces are more likely to be associated with contextal information than voices (Hanley et al 1998 and Chapter 3).

However, what is central to the research aims of this chapter is whether R responses continue to be predominantly associated with faces even when face recognition is reduced? To answer this question, an additional face condition was designed in order to reduce the overall recognition level of faces to a comparable level with voices. As previously pointed out, this technique was successful. However, despite the two conditions being quantitatively matched for overall levels of performance, the two conditions exerted qualitatively different levels of subjective awareness.

The analysis on the R and K data for blurred faces and voices showed that R responses were more strongly associated with blurred faces than voices, whereas K responses were more strongly associated with voices than blurred faces. This pattern of results fits in with the hypotheses set out in the introduction. Furthermore, the fact that R responses continued to be more strongly associated with faces than voices even when performance between the two measures was matched adds weight to the idea that faces are better than voices because the former modality is more likely to be based on contextual memory (remembering).

Experiment 4a, had no a priori assumptions to make in relation to differences in guess responses across the two modalities. As mentioned previously, guess responses were added to the standard R/K paradigm to prevent participants from contaminating K responses with non-memory strategic judgments (Gardiner et al, 2002). However, what immediately became apparent with the inspection of figure 6.1.a, is that guess responses were higher in the voice condition than in the blurred face condition. This observation was confirmed statistically. This finding, although not initially relevant to the research aims of the study, suggests an alternative interpretation of the data. The current interpretation of the data is that blurred faces and voices are weighted differently in terms of remembering and knowing. However, this significant difference in guess responses suggests that the two modalities do not differ qualitatively from each other. Instead they differ quantitatively, with the high proportion of R responses in the blurred face condition occuring as the result of greater confidence associated with blurred faces than voices (Donladson, 1996).

This quantitative account of the R and K data is a recurring theme in the R/K literature. It is often approached by asking participants to follow-up their 'yes' responses with a confidence judgement instead of an R, K or G judgement. Experiment 4b, used confidence judgements to assess this alternative interpretation of the data. If the R and K data is consistent with the trace-strength account, then blurred faces should be associated with higher levels of confidence than voices. However the analysis from Experiment 4b was not consistent with this idea. As illustrated in figure 6.2.a , there is no evidence that blurred faces produced higher levels of confidence than voices. This outcome supports the earlier findings by Mäntylä (1997) who showed that levels of subjective awrareness are not related to the levels of confidence associated with people's recognition judgements.

The only area of concern with the data from Expriment 4b, was that it failed to replicate equal levels of performance between blurred faces and voices. Whilst the two modalities were matched for overall hit rate, they differed significantly in their false alarm rate – a difference which was maintained for A' measures as well. Therefore, the R/K/G judgments were re-instated in experiment 4c. This experiment was a replication of experiment 4a with the following modifications. As the central objective of this chapter was to examine selective effects on R and K judgements when overall performance is matched, participants were only tested with blurred faces and voices. Furthermore, in order to reduce the possibility that the recognition of blurred faces is easier than voices, an additional level of blurring was applied to the faces to make them more difficult to recognize.

Increasing the level of blur from 13 in Experiment 4a to 26.25 in Experiment 4c, successfully decreased the overall levels performance in the blurred condition. Although this level of degradation produced recognition rates that were significantly lower than Experiment 4a, critically the overall recognition level did not differ significantly from the voice condition (see table 6.3.a).

Objective (table 6.3.c) and subjective assesments (table 6.3.e) of R, K and G responses relpicated the findings observed in experiment 4a, showing that participants used these memory judgements in a consistent way. In experiment 4c, all levels of subjective awareness were compared across the two modalities. In contrast to experiment 4a, guess responses did not significantly differ across the two modalities. Such an outcome is not consistent with the view that blurred faces are easier to recognize than voices. However, consistent with Experiment 4a, blurred faces were again associated with higher levels of remembering than voices. Voices on the other hand, showed higher levels of knowing than blurred faces.

The results from experiments 4c are more robust than those obtained from 4a. This is mainly beacuse the differences in the guess responses was eliminated in experiment 4c, therefore strengthening the view that faces and voices entail qualitatively different memory processes.

This dissociation of remember and know responses between blurred faces and voices adds to the previous research that has observed such dissociations across different materials (Gardiner and Java, 1990, Rajaram et al, 2002) and populations (Perfect et al, 1995). However, what are the theoretical underpinnings for the dissociation between blurred faces and voices?

Some of the answers may lie in the accounts that have been provided for the dissociation between remember and know responses for words and non-words. For example, Gardiner and Java (1990) explained the interaction of words over non-words in terms of the type of rehearsal strategy that was used. The finding that non-words tended to be more strongly associated with knowing than remembering was explained in terms of shallow or maintenance rehearsal. Words on the other hand produced more remember than know responses and this was taken to be a by-product of elaborative rehearsal. How can this rehearsal account be used to explain the differences between faces and voices and why might such differences occur in the first place?

In Brandt et al's study (2003) remember responses are more strongly associated with newly learnt distinctive faces than typical faces. This led the authors to argue that distinctive faces produce more remember responses because they are elaborately encoded. This is because distinctive faces stand out from typical faces and a possible consequence of this is that distinctive faces may receive additional attentional processing relative to typical faces.

Evidence to support this argument was obtained when participants were asked to study the distinctive and typical faces under more attentively demanding conditions. When participants were asked to count backwards whilst studying the faces, there were fewer remember responses made to distinctive faces compared to the control condition where participants were not given this additional task. When elaborate encoding is prevented, then the chances that recognition will be accompanied by remembering is greatly reduced. Therefore one possible explanation that could be proposed is that voices, compared to faces, are not elaborately encoded.

One way to investigate this hypothesis is to look at how subjective awareness contributes to the recognition of newly learnt voices. The results from this study have shown that famous voices were not as strongly associated with remembering compared to blurred faces. Whether this is also the case when new voices are learnt would be an interesting avenue to pursue for future research.

It is also important to consider that faces and voices co-occur extremely reliably. In our general day-to-day social encounters, we get to see the person's face and hear the person's voice when we are interacting with them. Because we have two very salient social cues available to us, it could be possible that there is competition for attentional resources when it comes to the encoding of the social encounter. If faces capture the majority of the attentional resources available, then in accordance with Brandt et al's findings, this would lead to more elaborate rehearsal and subsequent remembering. If this is the case then this would mean that voices would not be as elaborately processed as faces, which could reduce the level of remember responses. The study by McAllister, Dale, Bregman, McCabe and Cotton (1993) does seem to provide some support for such an account.

These authors noted that voice identification performance was worse for participants who were given a photograph of the speaker to view, than for participants who were only exposed to the voice without seeing the face. Such a pattern in performance does seem to suggest that visual and auditory cues compete for attentional information, with visual information capturing the majority of the resources available. Although this study by McAllister et al (1993) did not use the standard R/K procedure, it would be interesting to adopt this procedure when looking at how newly learnt faces and voices are recognized. One area of interest could look at the contribution of remembering and knowing when voices are studied in the absence of facial input compared to when they are studied alongside facial information.

One issue that is directly related to this competition for attentional resources hypothesis is how salient are these social cues to the recipient and whether this level of saliency is the same for both modalities? Starting with voices, research has consistently shown that variations in pitch and speech rate convey a person's identity effectively (Schweinberger, 2001, Schweinberger et al, 1997, Van Lancker, Kreiman and Wickens, 1985, Van Lancker, Kreiman and Emmorey, 1985). However, what needs to be measured is how these perceptual cues are encoded and subsequently recognized. The R/K/G paradigm can be used in such perceptual studies to elucidate the memory processes involved when these cues are recognized.

In terms of facial cues, research has already begun to adopt the R/K/G paradigm in an attempt to understand which facial cues are encoded and used for recognition. For example, Wright and Sladden (2003) recently showed that encoding a person's hair is critical for producing a remember response. This observation by Wright and Sladden (2003) may shed some light on the findings from the blurred face condition in experiment 4a and 4c. Specifically, participants may have used these cues (which are relatively preserved in blurring) to help them with the recognition process. So even though blurred faces and voices were matched in terms of overall recognition, they may not have been matched in terms of the saliency of the cues available in each modality. It may be the case that hair cues along with other physical facial attributes could have promoted Remember responses in the blurred face condition. This leads us to the question of whether voice recognition difficulties occur at the stage of encoding or at the stage of retrieval.

The research findings from Hanley et al's study (1998) have shown that compared to faces, voices are relatively problematic when it comes to the retrieval of person-specific information. This finding was observed consistently in Chapter 3 of this thesis. If the acoustic properties of a voice (e.g; pitch and speech rate) can be used to convey identity information, why then, are they not as salient as the cues available in a face? One possibility is that faces and voices are associated with different inherent perceptual properties which may promote or delay the encoding process. This idea has been elegantly articulated by Olsson, Juslin and Winman (1998), who stated that 'the defining characteristics of a voice are extended in time, whereas all information is available instantly when we look at a face' (page 116). This in turn would suggest that faces are slightly ahead of voices in terms of the time course for the encoding process.

Another interesting aspect of the data relates to the dual-component view of Remembering and Knowing. According to Gardiner et al (2002), Know responses entail a qualitatively different form of memory that can be used alongside Remembering to improve overall recognition. This hypothesis has been used to counter Donaldson's (1996) view that the two memory processes differ simply in terms of subjective confidence. Gardiner et al (2002) provided support for their dual-component view by analysing the A' values for Remember responses versus Remember plus Know responses combined. Their results showed that when Knowing was added to Remembering, recognition (A') improved significantly compared to when recognition was based purely on Remember responses. A dual-component pattern was found in the standard face condition in experiment 4a. The analyses showed that the standard face condition was the only modality in which knowing significantly improved recognition above remembering alone (see table 6.1.d).

The R/K paradigm provides the researcher with a window on what participants are actually experiencing when they attempt to recognize a previously encountered item. Remembering was originally conceptualized to reflect episodic memory processes (Tulving, 1985), which appears to involve the integration of 'what', 'where' and 'when' something happened (Aggleton and Pearce, 2001). Recognition can also be achieved without contextually re-experiencing the event, but through simply finding the item familiar or 'knowing it'. This form of recognition was thought to reflect the properties of the semantic system. As a result, these phenomenal aspects of the R/K paradigm have often been termed as the 'first-person' approach to recognition memory (Gardiner, 2001).

The R/K paradigm in this study was applied to the recognition of faces and voices in the anticipation that it can provide a window on the breakdown of person recognition memory. One of the aims of this chapter was to gain a better understanding of why voices are less likely to be associated with contextual information than faces (Chapter, 3). To take account for any differences in overall levels of performance, standard faces were presented in a blurred format – a condition which produced overall levels of recognition similar to voices. Even under these circumstances, it was shown that voices behave very differently from faces. One explanation that emerged to account for the occupational retrieval advantage of faces over voices (Chapter 3) is that voice recognition, relative to face recognition, is more likely to be associated with knowing - a recognition route which is not based on contextual retrieval. These qualitative memory differences between faces and voices can not simply be explained away in terms of different levels of confidence between the two modalities (experiment 4b). Differences in faces and voices may begin at the encoding level, with the temporal properties of a voice and/or lack of perceptual cues impairing the encoding process relative to faces. This may subsequently affect recognition performance, where these perceptual cues are more salient for faces than voices.

# Chapter 7: Conclusions.

## 7.1: Conclusions.

This thesis aimed to investigate the memory processes involved in familiar voice recognition. The theoretical grounding for the experiments conducted in this thesis primarily stemmed from the research conducted on face recognition. Specifically, it aimed to elucidate the possible psychological factors that may underlie cross-modal differences in person-specific information. This issue was approached through a number of different perspectives. Therefore, this chapter is divided into four sections, each summarizing and reflecting the way in which these different approaches have contributed to the understanding of familiar voice recognition.

### Chapter 3: The relationship between familiarity and person identification. Is it the same for faces and voices?

The aim of the experiments carried out in Chapter 3 was to examine the familiar-only bias in voice recognition in more detail and by doing so, expand the study of person recognition errors to the auditory domain. There are two alternative ways of assessing the familiar-only bias in voices, both of which involve matching the levels of familiarity between faces and voices. One technique involves reducing the overall level of recognition in the face condition. This technique was adopted by Hanley et al (2000), who demonstrated that by blurring faces, the ability to retrieve occupational information was the same as voices. Under this particular manipulation, the familiar-only bias for voices was removed.

A less artificial way to equate familiarity across modalities is to ask the same participants to make familiarity judgments to both faces and voices. The fact that Hanley et al's (1998) study employed a between-subjects design meant that differences between the two modalities could have arisen due to differences in criteria adopted by the participants in the face and voice conditions. The hypothesis of interest was whether asking participants to make familiarity judgments to faces would enable participants to be more accurate in assigning their familiarity judgments to voices. This calibration hypothesis was adopted from the confidence-accuracy literature which showed that participants can use their confidence judgments on one task as a context for their judgment on another task. This has the overall effect of improving the confidence-accuracy relationship and is the most effective when items are 'mixed' together (Hollins and Perfect, 1997).

Experiment 1a, aimed to replicate Hanley et al's (1998) original findings using a between-subjects design. The results were largely consistent with Hanley et al's results and replicated the familiar only bias for voices with a new set of stimuli. Experiment 1b, attempted to eliminate criteria differences between faces and voices by asking the same participants to make familiarity judgments to both faces and voices. This change in design meant that familiarity decisions to faces should be on a comparable scale to voices. Even with a within-subjects design, faces deemed familiar elicited significantly more occupational details than voices deemed familiar.

Experiment 1c, aimed to test this 'mixing' effect in more detail. One possible explanation why the mixing effect failed to remove the familiar-only error in experiment 1b was that the opportunity for calibration between familiarity judgments for faces and voices was not optimal. In Experiment 1b, participants alternated their familiarity judgments between faces and voices after every fourth trial. Experiment 1c, examined whether alternating between faces and voices on every trial would encourage the participants to use the familiarity scale even more consistently. Even under this manipulation, the familiar-only bias for voices continued.

Throughout Experiments 1a –1c it became increasing apparent that the overall levels of recognition for the voice condition were worryingly low. At de-briefing participants often claimed that they found it difficult to discriminate between the American voices for famous and non-famous items. This observation was confirmed statistically. This advantage of British voices over American ones may reflect an auditory version of the other-race effect (Malpass and Kravitz, 1969). Therefore, a new selection of stimuli were selected, with the main criteria for selection being that the nationality of the stimuli matched the nationality of the participants. This new selection of stimuli were used in Experiments 1d and 1e.

Experiments 1d, used a between-subjects design with this new set of stimuli and replicated the familiar-only bias for voices observed in Experiment 1a. Experiment 1e, used a within-subjects design in which participants were required to make familiarity decisions to faces and voices on every trial. Again, the retrieval of person-specific information remained particularly elusive for voices.

The concluding arguments from Chapter 3 are as follows. If the familiar-only bias in Hanley et al's (1998) experiment resulted in terms of poor calibration between faces and voices, then any differences between the two modalities should have disappeared in circumstances in which participants are provided with an opportunity to use the same familiarity judgments for faces and voices (Experiments 1b, 1c and 1e). The fact that the familiar-only bias for voices remains across the five experiments reported in Chapter 3 suggests that a calibration problem may not be a suitable explanation of the differences between the modalities.

It appears that participants mean something different by 'very familiar' when applied to voices relative to what it means when applied to faces. What this series of experiments have shown is just how difficult it is to change this bias. Unlike differences in confidence (Hollins and Perfect, 1997), familiarity decisions can not be recalibrated by employing a within-subjects design.

### Chapter 4: Confidence judgments in person recognition memory.

Experimental Chapter 4 aimed to extend the findings obtained in Chapter 3 by asking participants to make confidence judgments to faces and voices. Previous work from the confidence-accuracy (C-A) literature has shown that participants' overall recognition for newly learnt voices is considerably lower than that for newly learnt faces even at the highest confidence category (Olsson et al, 1998). This is referred to in the C-A as the over-confidence bias.

Experiment 2 in Chapter 4 studied this overconfidence bias by asking participants to make confidence judgments to famous faces and voices. The results extended the findings from Olsson et al's (1998) study by showing that face recognition reached near perfect calibration (0.99) for 'sure' responses which was the highest confidence category in this experiment. Recognition of famous voices for the same category was 0.85, a value which significantly differed from the face condition.

The next stage of the analysis assessed whether judgments of confidence behave similarly to judgments of familiarity when participants are required to retrieve person specific identity information to the faces and voices that they recognize. The previous chapter showed that with familiarity judgments, faces deemed familiar were much more likely to be associated with biographical information than voices. However, when participants were required to make confidence judgments, the analyses failed to reveal any modality differences for occupational and name recall. In terms of person-specific information, these results suggest that judgments of familiarity can be dissociated from judgments of confidence. What is needed to further validate this conclusion in the future, is a study that uses the same response scale for familiarity and confidence judgments.

This chapter also used a qualitative approach to studying face and voice recognition. Specifically, Yonelinas' (1999) ROCs and z-ROCs approach was used in an attempt to gain a better understanding of the recognition processes that govern face and voice recognition. The initial analyses of the ROCs and z-ROCs produced qualitatively different curves for the face and voice conditions. The first impression from the face condition was that a recollective component as well as a familiarity component was contributing to the recognition process. The contribution of recollection and familiarity is consistent with Yonelinas' (1999) dual-component recognition model. However, detailed analyses of the z-ROC for the face condition ruled out the possibility of a recollective component. Instead it was tentatively concluded that the face condition elicited similar patterns of performance as predicted by the unequal variance familiarity model.

The unequal variance familiarity model was also used to account for the data in the voice condition. The ROCs and z-ROCs approach suggests that both modalities entail recognition that is based on assessments of familiarity with the target item. However, a number of different problems emerged with the ROCs and z-ROCs which made the conclusions from this experiment tentative at best. Firstly, the ceiling effect in the face condition could have masked the contributions of a recollective component. Furthermore, the z-ROCs from the voice condition produced a curve that was not consistent with Yonelinas' observations. Such deviations in curve linearity have been observed by other researchers (Ratcliff, McKoon and Tindall, 1994, Glanzer, Hilford, Kim and Adams, 1999a, Rotello, Macmillan and Van Tassel, 2000) and have been used to put forward the idea that a third component, along with a familiarity and recollection process may be operating in recognition judgments.

With this in mind, the remainder of the thesis continued to assess whether faces and voices differed qualitatively by using a different approach which explicitly allowed for a guessing component to be measured. The technique of interest was the Remember / Know approach (Tulving, 1985, Gardiner, 1988). Before this experiment was carried out, a face condition that produced similar overall levels of recognition as the voice condition was needed. This was the central aim of the pilot study conducted in Chapter 5.

### Chapter 5: Matching face and voice recognition.

The pilot study conducted in this chapter aimed to create a face condition which produced similar levels of hit and false alarm rates to voices. This was accomplished by adopting three visual degradation techniques which have been extensively used in the face recognition literature. The three degradation conditions included inverted faces (Yin, 1969, Lander, Christie and Bruce, 1999), blurred faces (Collishaw and Hole, 2000, Hanley and Turner, 2000, Lander, Bruce and Hill, 2001) and photographic negatives (Knight and Johnston, 1997).

The results showed that the blurred face condition was the most effective technique in producing a facial condition that was similar to the voice condition both in terms of overall recognition and knowledge. The analyses extended the work of Experiment 2 by examining the ROCs and z-ROCs produced by the different degradation techniques. The previous chapter highlighted the problems of ceiling effects in the interpretation of ROCs and z-ROCs for the face condition (Yonelinas, 2001). When this was removed by applying a degradation technique to the faces, the blurred and inverted conditions continued to produce ROCs and z-ROCs that were consistent with the unequal variance familiarity model. Negative faces on the other hand, showed evidence of a recall component, thus supporting Yonelinas' (1999) dual-component theory.

It was concluded that when measures are taken to remove ceiling effects in face recognition, then the contribution of a recollective process as generated through ROCs and z-ROCs is possible to pick out. One hypothesis that was proposed to account for the recollective component in negative faces was that this particular technique preserves both configural and featural cues (Collishaw and Hole, 2000) and that both of these cues need to be available in order to promote recollection.

### Chapter 6: The R and K paradigm. Can it be applied to the recognition of familiar faces and voices?

The final series of experiments in Chapter 6 adopted the 'first person' approach (Gardiner, 2001) to assess the differences in face and voice recognition by using the Remember / Know (Tulving, 1985, Gardiner, 1988) paradigm. Both objective and subjective assessments of participants' remember, know and guess responses showed that participants understood the instructions and used them in a valid way (Gardiner et al, 2002).

The critical comparison of interest was between the blurred face and voice conditions. By presenting faces in a blurred format, overall recognition was reduced to the same level as the voice condition. The key aspect of the analysis showed that despite the two modalities being matched in terms of overall recognition, the two modalities produced selective effects on remembering and knowing. Specifically, blurred faces were much more likely to be associated with remembering than voices, whereas voices produced a higher proportion of know responses than blurred faces.

The hypothesis that the two modalities simply differed in terms of subjective confidence (Donaldson, 1996) was also tested in this chapter and discounted. This particular outcome supports previous research with novel faces which showed that remember responses are not simply a product of trace strength (Mäntylä, 1997).

Blurred faces and voices continued to produce selective effects on remembering and knowing even when the recognition level in the blurred faces was made much harder. One hypothesis that was put forward to account for this cross modal difference is that faces convey salient perceptual cues which may encourage better encoding of the face and which in turn can be used to aid recognition (Wright and Sladden, 2003). Nevertheless, the key conclusion from this study is that the advantage of faces over voices in terms of person-specific information may be understood in terms of differences in memory processes. That is, faces are more likely than voices to engage in a memory process that is governed by contextual information. This qualitative difference persists even when differences in overall recognition have been taken into account.

## 7.2: Final Comments.

The validity of person recognition memory models (Bruce and Young, 1986, Burton, Bruce and Johnston, 1990) has largely been based on the research that has been conducted on face recognition. Although these models posit an auditory route alongside the visual one, very little is known how effectively this route operates. The aim of this thesis was to provide a better understanding of how person-specific information about familiar voices is retrieved by comparing face and voice recognition within the same series of experiments.

Using a selection of famous and non-famous faces and voices, this issue was tackled from three different yet complimentary perspectives; familiarity judgments, confidence judgments ROCs and z-ROCs and the R/K paradigm. The consistent findings across the experiments (albeit Experiment 2) is that relative to faces that are found familiar, the retrieval of person-specific information was particularly problematic for voices. Considering the architectural properties of these models, this problem may reflect weak associative links from the VRU to the PINs than the FRU and the PINs (Hanley and Turner, 2000).

Perhaps the most theoretically interesting finding came from the use of R/K paradigm, which suggests that the recognition of pre-experimentally familiar faces and voices may rely on qualitatively different memory processes. This approach is currently being adopted in the face recognition domain (Mäntylä, 1997, Brandt et al, 2003, Wright and Sladden, 2003) and has provided a number of interesting findings on the encoding and recognition of faces. It is time that voice recognition research catches up with the developments made in face recognition. Applying the R/K paradigm to study how voices are encoded and recognized is one potential way in which the gap can be reduced.

# References.

Aaronson, D. and Watts, B. (1987). Extensions of Grier s computational formulas for A' and B to below-chance performance. *Psychological Bulletin, 102,* 439 — 442.

Aggleton J. P. and Pearce J. M. (2001). Neural systems underlying episodic memory: insights from animal research. *Philosophical Transactions: Biological Sciences, 356,* 1467-1482.

Baddeley, A. D. (1997). *Human Memory: Theory and Practice.* Hove: Psychology Press Ltd.

Ball, M. J. and Rahilly, J. (1999). *Phonetics: The science of speech.* London: Arnold.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P. and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403,* 309 — 312.

Belin, P., Fecteau, S. and B dard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129 — 135.

Boles, D. B. (1984). Global versus local processing: is there a hemispheric dichotomy. *Neuropsychologia, 22,* 445 — 455.

Bothwell, R.K., Brigham, J.C. and Deffenbacher (1987). Correlation of Eyewitness Accuracy and Confidence: Optimality Hypothesis Revisited. *Journal of Applied Psychology, 72,* 691 — 695.

Brandt, K., Macrae, N. C., Schloerscheidt, A. M and Milne, A. B. (2003). Remembering or knowing others? Person recognition and recollective experience. *Memory, 11,* 89-100.

Brennen, T., Baguley, T., Bright, J. and Bruce, V. (1990). Resolving semantically induced tip-of-the-tongue states for proper nouns. *Memory and Cognition, 18,* 339 — 347.

Brennen, T., David, D., Fluchaire, I. and Pellat, J. (1996). Naming faces and objects without comprehension: A case study. *Cognitive Neuropsychology, 13,* 93 — 110.

Brown, R. and McNeill, D. (1966). The 'tip of the tongue' phenomenon. J*ournal of Verbal Learning and Verbal Behaviour, 5,* 325 — 337.

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

Bruce, V. and Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology, 76,* 373 — 383.

Bruce, V. and Valentine, T. (1986). Semantic priming of familiar faces. *Quarterly Journal of Experimental Psychology, 38A,* 125 — 150.

Bull, R., Rathborn, H. and Clifford, B. R. (1983). The voice-recognition accuracy of blind listeners. *Perception, 12,* 223 — 226.

Burton, A. M. and Bruce, V. G. (1992). I recognize your face but I can't remember your name: A simple explanation. *British Journal of Psychology, 81,* 361 — 380.

Burton, A. M. and Bruce, V. G. (1993). Naming faces and naming names: Exploring an interactive activation model of person recognition. *Memory, 1,* 457 — 480.

Burton, A. M., Bruce, V. G. and Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81,* 361 — 380.

Burton, M. A., Bruce, V. and Hancock, P. J. B. (1999). From Pixels to People: A Model of Familiar Face Recognition. *Cognitive Science, 23,* 1 — 31.

Bushnell, I. W., Sai, F. and Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology, 7,* 3 — 15.

Cobb, N., Lawrence, D. and Nelson, N. (1979). Report on blind subjects' tactile and auditory recognition for environmental stimuli. *Perception and Motor Skills, 48,* 363 — 366.

Cohen, G. (1990b). Why is it difficult to put names to faces. *British Journal of Psychology, 81, 287 — 297.*

Coleman, R. O. (1976). A comparison of the contributions of two voice quality characteristics to the perception of male-ness and female-ness in the voice. *Journal of Speech and Hearing Research, 19,* 168 - 180.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour 60,* 773 — 780.

Collishaw, S.M. and Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29,* 893 — 909.

Cook, S. and Wilding, J. (1997b). Earwitness testimony 2: Voices, Faces and context. *Applied Cognitive Psychology, 11,* 527 — 541.

Costermans J, Lories G and Ansay, C. (1992). Confidence Level and Feeling of Knowing in question Answering: The weight of Inferential Processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 142 — 150.

Dalla Barba, G. (1993). Confabulation: Knowledge and recollective experience. *Cognitive Neuropsychology, 10,* 1-20.

Damasio, A. R., Damasio, H. and Van Hoesen, G. W. (1992). Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology, 32,* 89 — 109.

de Haan, E. H. F., Young, A. W. and Newcombe, F. (1991). A dissociation between the sense of familiarity and access to semantic information concerning familiar people. *European Journal of Cognitive Psychology, 3,* 51 — 67.

DeCasper, A.J and Fifer, W.P. (1980). Of Human bonding: Newborns prefer their Mothers' voices. *Science, 208,* 1174 — 1176.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory and Cognition, 24,* 523 — 533.

Ekman, P. and Friesen, W. V. (1975). *Pictures of Facial Affect.* Palo Alto, CA: Consulting Psychologists Press.

Eldridge, L. L., Sarfatti, S. and Knowlton, B. J. (2002).The effect of testing procedure on remember-know judgments. *Psychonomic-Bulletin-and-Review, 9,* 139 —145.

Ellis, A. W., Young, A. W. and Flude, B. M. (1990). Repetition priming and face processing: Priming occurs within the system that responds to the identity of a face. *Quarterly Journal of Experimental Psychology, 39A,* 193 — 210.

Ellis, H. D. (1983). The role of the right hemisphere in face perception. In Young, A. W (Eds). *Functions of the right cerebral hemisphere (pp 33 —44).* London: Academic Press.

Ellis, H. D., Jones, D. M. and Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88,* 143 — 156.

Flude, B. M., Ellis, A. W. and Kay, J. (1989). Face processing and name retrieval in an anomic aphasic. *Brain and Cognition, 11,* 60 — 72.

Froming, K. B., Levy, M. and Ekamn, P. (2003). Comprehensive Affect Testing System.

Gardiner, J. M (2001). Episodic memory and autonoetic consciousness: A first-person approach. In Baddeley, A., Aggleton, J. P. and Conway, M. (Eds), *Episodic memory: New directions in research (pp. 11 — 30).*Oxford: Oxford University Press.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory and Cognition, 16,* 309-313.

Gardiner, J. M., Kaminska, Z., Dixon, M. and Java, R. I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin and Review, 3,* 366-371.

Gardiner, J. M., Ramponi, C. and Richardson — Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know and guess responses. *Memory, 10*, 83 — 98.

Gardiner, J. M., Ramponi, C. and Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition, 7,* 1-26.

Gardiner, J.M., and Java, R.I. (1990). Recollective experience in word and nonword recognition. *Memory and Cognition, 18,* 23 — 30.

Glanzer, M., Hilford, A., Kim, K. and Adams, J. K. (1999a). Further tests of dual-process theory: A reply to Yonelinas (1999a). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 522 — 523.

Glanzer, M., Hilford, A., Kim, K. and Adams, J. K. (1999b). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 500-513.

Goggin, J. P., Thompson, C. P., Strube, G. and Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19,* 448 — 458.

Hanley, J. R. and Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A,* 1105 -1116.

Hanley, J. R. (1995). Are names difficult to recall because they are unique? A case study of a patient with anomia. *Quarterly Journal of Experimental Psychology, 48A,* 487 — 506.

Hanley, J.R. and Cowell, E.S. (1988). The effects of different types of retrieval cues on the recall of names of famous faces. *Memory and Cognition, 16,* 545 — 555.

Hanley, J.R., Smith, S.T., and Hadfield, J. (1998). I recognize you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *The Quarterly Journal of Experimental Psychology, 51A,* 179-195.

Haslam, C., Cook, M. and Coltheart, M. (2001). 'I know your name but not your face': Explaining modality-based differences in access to biographical knowledge in a patient with retrograde amnesia. *Neurocase, 7,* 189-199.

Hauser, M. (1996). *The Evolution of Communication.* Cambridge: Mass, MIT Press.

Hay, D. C., Young, A. W. and Ellis, A. W. (1991). Routes through the face recognition system. *Quarterly Journal of Experimental Psychology, 43A,* 761 — 791.

Heathcote, A. (2003). Item Recognition Memory and the Receiver Operating Characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1210 — 1230.

Hodges, J. R. and Greene, J. D. W. (1998). Knowing about people and naming them: Can Alzheimer's disease patients do one without the other? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51A,* 121-134.

Hollins, T. S. and Perfect, T. J. (1997). The confidence-accuracy relation in eyewitness event memory: The mixed question type effect. *Legal and Criminological Psychology, 2,* 205 —218.

Humphreys, G. W., Donnelly, N. and Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologia, 31,* 173 — 181.

Insley, S. J. (2000). Long-term vocal recognition in the northern fur seal. *Nature, 406,* 404-405.

James, W. (1890). *Principles of psychology.* New York: Holt.

Johnson, M. H. (1992). Imprinting and the development of face recognition: From chick to man. *Current Directions in Psychological Science, 1,* 52-55.

Johnston, A., Hill, H. and Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception, 21,* 365 — 375.

Juslin, P., Olsson, N. and Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 1 — 13.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3,* 759 — 763.

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

Kisilevsky, B.S., Hains, S. M. J., Lee, K., Xie, X., Huang, H. H., Hui, Y. H., Zhang, K. and Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science, 14,* 220-224.

Knight, B. and Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4,* 265 — 273.

Ladefoged, P. (2001). *A course in phonetics.* Fort Worth: Harcourt College Publishers.

Ladefoged, P. and Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics, 49,* 43 — 51.

Lander, K., Bruce, V. and Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology, 15,* 101-116.

Lander, K., Christie, F. and Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition, 27,* 974 — 985.

Legge, G. E., Grosmann, C. and Pieper, C. M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 298 — 303.

Liu, C. H. and Chaudri, A. (2003). What determines whether faces are special? *Visual Cognition, 10,* 385 — 408.

Macrae, C. N. and Lewis, H. L. (2002). Do I know you? Processing Orientation and Face Recognition. *Psychological Science, 13,* 194 — 196.

Malone , D.R., Morris, H.H., Kay, M.C. and Levin, H. S. (1982). Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery and Psychiatry, 45,* 820 - 822.

Malpass, R.S. and Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330-334.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252-271.

Mann, V. A., Diamond, R. and Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27,* 153 — 165.

M ntyl , T. (1997). Recollection of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1203 — 1216.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Maylor, E. A. (1995). Remembering versus knowing television theme tunes in middle-aged and elderly adults. *British Journal of Psychology, 86,* 21 — 25.

McAllister, H. A., Dale, R. H. I., Bregman, N. J., McCabe, A. and Cotton, C. R. (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and applied Social Psychology, 14,* 161 — 170.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17,* 249 — 271.

McGehee, F. (1944). An experimental study of voice recognition. *Journal of General Psychology, 31,* 56 — 65.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746 — 748.

McLelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. *Psychological Review, 88,* 375 — 407.

McWeeny, K. H., Young, A. W., Hay, D. C., Ellis, A. W. (1987). Putting names to faces. *British Journal of Psychology, 78,* 143 — 149.

Meuwly, D. (2000). Voice Analysis. In Siegel, J., Saukko, P. and Knupfer, G (Eds). *Encyclopedia of Forensic Science* (pp. 1413 — 1420). London: Academic Press.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In Kolers, P.A., Wrolstad, M. and Bouma, H. (Eds), *Processing of visible language.* New York: Plenum.

Neuner, F. and Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition, 44,* 342-366.

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology, 4,* 504 — 511.

Olsson, N., Juslin, P and Winman, A. (1998). Realism of Confidence in Earwitness Versus Eyewitness Identification. *Journal of Experimental Psychology: Applied, 4,* 101 — 118.

Ormerod, D. (2001). Sounds Familiar? — Voice Identification Evidence*Criminal Law Review,* 595 — 622.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C. and Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development, 18,* 79 — 85.

Perfect, T. J., Williams, R. B. and Anderton-Brown, C. (1995). Age differences in reported recollective experience are due to encoding effects, not response bias. *Memory, 3,* 169-186.

Perfect, T.J, Hollins, T.S. and Hunt, A.L.R. (2000). Practice and feedback effects on the confidence-accuracy relation in and feedback effects on the confidence-accuracy relation in eyewitness memory. *Memory, 8,* 235 — 244.

Perfect, T.J, Watson, E.L. and Wagstaff, G.F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology, 78,* 144-147.

Pilling, M., Wiggett, A., zgen, E. and Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory and Cognition, 31,* 538 — 551.

Pinker, S. (1994). *The language instinct: the new science of language and mind.* London: Allen Lane.

Ptacek, P. H., and Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research, 9,* 273 — 277.

R v R (2000). *Criminal Law Review,* 183 — 185.

R v. Robb (1991). *Criminal Appeal Review, 161.*

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 365 — 377.

Rajaram, S., Hamilton, M. and Bolton, A. (2002). Distinguishing states of awareness from confidence during retrieval: Evidence from amnesia. *Cognitive, Affective and Behavioral Neuroscience, 2,* 227 — 235.

Rapcsak, S. Z., Polster, M. R., Comer, J. F. and Rubens, A. B. (1994). False recognition and misidentification of faces following right hemisphere damage. *Cortex, 30,* 565 — 583.

Ratcliff, R., McKoon, G. and Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 763 — 785.

Roberson, D. and Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory and Cognition, 28,* 977-986.

Rotello, C. M., Macmillan, N.A. and Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67 — 88.

Sacks, O. W. (1986). *The man who mistook his wife for a hat.* London: Picador.

Sapir, E. (1927). Speech as a personality trait. *American Journal of Sociology, 32,* 892 — 905.

Saslove, H. and Yarmey, A.D. (1980). Long-term Auditory Memory: Speaker Identification. *Journal of Applied Psychology, 65,* 111 — 116.

Scanlan, L. C. and Johnston, R. A. (1997). I recognize your face but I can't remember your name: A grown-up explanation. *Quarterly Journal of Experimental Psychology, 50A,* 183 — 198.

Scherer, K. R. (1981). Speech and emotional states. In Darby, J.K. (Eds). *Speech Evaluation in Psychiatry. (pp 189 — 221).* New York: Grune and Stratton, Inc.

Schmidt-Nielsen, A. and Stern, K.R. (1985). Identification of known voices as a function of familiarity and narrow-band coding. *Journal of the Acoustical Society of America, 77,* 658 — 663.

Schooler, J.W. and Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22,* 36 — 71.

Schwartz, M. F. and Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America, 44,* 1736 — 1737.

Schweich, M., van der Linden, M., Br dart, S., Bruyer, R., Nelles, B. and Schils, J. P. (1992). Daily-life difficulties in person recognition reported by young and elderly subjects. *Applied Cognitive Psychology, 6,* 161 — 172.

Schweinberger, S. R. (2001). Human brain potential correlates of voice priming and voice recognition. *Neuropsychologia, 39,* 921-936.

Schweinberger, S. R., Herholz, A. and Stief, V. (1997). Auditory long-term memory: Repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A,* 498-517.

Schweinberger, S.R., Herholz, A., Sommer, W. (1997). Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *Journal of Speech and Hearing Research, 40,* 453-463.

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

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J. and Fink, G. R. (2001).The neural correlates of person familiarity: A functional magnetic resonance imaging study with clinical implications. *Brain, 124,* 804 — 815.

Shipp, T. and Hollien, H. (1969). Perception of the aging male voice. *Journal of Speech and Hearing Research, 12,* 703 —710.

Smith, V. L., Kassin, S.M. and Ellsworth, P.C. (1989). Eyewitness accuracy and confidence: within versus between-subjects correlations. *Journal of Applied Psychology, 74,* 356 — 359.

Toates, F. (1998). Sensory Systems. In Eysenck, M. (Eds), *Psychology. An integrated approach* (100 — 138). New York: Addison Wesley Longman, Inc.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1-12.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471 — 491.

Valentine, T., Chiroro, P. and Dixon, R. (1995). An account of the own-race bias and the contact hypothesis on a 'face space' model of face recognition. In Valentine, T, (Eds). *Cognitive and Computational Aspects of Face Recognition. Explanations in face space (pp 69 — 92).*Guildford: Biddles, Ltd.

Van Lancker, D. and Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25,* 829 — 834.

Van Lancker, D. R., Cummings, J. L., Kreiman, J. and Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex, 24,* 195 — 209.

Van Lancker, D., Kreiman, J and Emmorey, K (1985). Familiar voice recognition: patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics, 13,* 19 — 38.

Van Lancker, D., Kreiman, J. and Wickens, T. D. (1985). Familiar voice recognition: Patterns and parameters: II. Recognition of rate-altered voices. *Journal of Phonetics, 13,* 39-52.

Voiers, W. D. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America, 36,* 1065 — 1073.

Ward, J., Parkin, A. J., Powell, G., Squires, E. J., Townshend, J. and Bradley, V. (1999). False recognition of unfamiliar people: "Seeing film stars everywhere." *Cognitive Neuropsychology, 16,* 293 — 315.

Waters, K. (1992). Modelling three-dimensional facial expressions. In V. Bruce and M. Burton (Eds). *Processing Images of Faces.* (pp 202-227). Norwood, NJ: Ablex Publishing Corporation.

Wilding, J., Cook, S. and Davis, J. (2000). Sound familiar? *Psychologist, 13,* 558 — 562.

Williams, M. and Smith, H. V. (1954). Mental disturbances in tuberculous meningitis. *Journal of Neurology, Neurosurgery and Psychiatry, 17,* 173 — 182.

Wright, D.B. and Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica, 114,* 101 — 114.

Yarmey, A. D. (1973). I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory and Cognition, 1,* 287 — 290.

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy and Law, 1,* 792 — 816.

Yarmey, A. D., Yarmey, A. L. and Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8*, 453 — 464.

Yarmey, A.D, Yarmey, A.L., Yarmey, M.J. and Parliament, L. (2001). Common sense beliefs and the identification of familiar voices. *Applied Cognitive Psychology, 15,* 283 - 299.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 41-145.

Yonelinas, A. P. (2001). Consciousness, Control, and Confidence: The 3 Cs of Recognition Memory. *Journal of Experimental Psychology: General, 130,* 361 — 379.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415 — 1434.

Yonelinas, A. P., Kroll, N.E.A., Dobbins, I. G. and Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgements. *Psychonomic Bulletin and Review, 6,* 654 — 661.

Yonelinas, A.P., Kroll, N. E. A., Dobbins, I., Lazzara, M. and Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12,* 323 — 339.

Young, A. W., Ellis, A. W. and Flude, B. M. (1988). Accessing stored information about familiar people. *Psychological Research, 50,* 111 — 115.

Young, A. W., Hay, D. C., and Ellis, A. W. (1985a). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *Quarterly Journal of Experimental Psychology, 38A,* 297 — 318.

Young, A. W., McWeeny, K. H., Hay, D. C. and Ellis, A. W. (1986). Access to identity-specific semantic codes from familiar faces. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38A,* 271-295.

Young, A.W., Mcweeny, K.H., Ellis, A.W. and Hay, D.C. (1986). Naming and categorizing faces and written names. *The Quarterly Journal of Experimental Psychology, 38A,* 297-318.

# APPENDIX 1.

List of celebrities used in Experiments 1a, 1b and 1c.

Ainsley Harriot
Andrew Lloyd Weber
Carol Voderman
Celine Dion
Chris O'Donnell
Clint Eastwood
David Ginola
Dawn French
Des O'Connor
Esther Rantzen
Goldie Hawn
Jennifer Lopez
Jo Brand
Joanna Lumley
John Travolta
Julia Roberts
Liam Neeson
Lord Parkinson
Lulu
Margaret Thatcher
Martine McCutheon
Michael Barrymore
Michael Heseltine
Michael Portillo
Noddy Holder
Pierce Brosnan
Robin Williams
Ronan Keating
Ruby Wax
Sally Gunnell
Tom Jones
Tommy Vance

## APPENDIX 2.

Donaldson’s (1996) (formula for sensitivity (A’):

0.5 + [ (Hits – False Alarms) (1 + Hits – False Alarms)] / [4 Hits (1 – False Alarms)]

For Bias (B”D):

[(1-Hits)(1-False Alarms) – (Hits)(False Alarms)] / [(1-Hits)(1-False Alarms) + (Hits)(False Alarms)]

When Hits < False Alarms, then the A’ formula by Aaronson and Watts (1987) was used:

A’ = 0.5 – (False Alarms – Hits)(1+False Alarms – Hits)/[4 False Alarms (1 – Hits)]

# APPENDIX 3.

List of celebrities used in Experiments 1d, 1e, 2, 3, 4a, 4b and 4c.

Anne Kirk Bride
Anne Robinson
Anne Widdecombe
Anthea Turner
Bob Monkhouse
Boy George
Bruce Forsyth
Carol Voderman
Chris Eubank
Chris Tarrant
David Beckham
Dawn French
Eammon Holmes
Elton John
Emma Bunton
Esther Rantzen
Gary Linekar
George Best
Jarvis Cocker
Joanna Lumley
John Major
Johnny Vegas
Julie Walters
Kate Winslet
Kirsty Young
Liam Neeson
Linford Christie
Lulu
Margaret Thatcher
Mark Lamarr
Michael Barrymore
Michael Heseltine
Michael Portillo
Noddy Holder
Ricky Tomlinson
Rod Stewart
Ronan Keating
Ronnie Barker
Ronnie Wood
Sally Gunnell
Sean Connorey
Sophie Ellis Bextor
Stephen Fry
Sting
Terry Wogan
Tom Jones
Ulrika Johnsson
Vinnie Jones

# APPENDIX 4.

## Instructions used in Chapter 6.

In this test you will see a series of faces[*], one at a time. Some of these faces belong to famous people and some are from non-famous people. The famous faces are selected from a wide range of fields including: politics, comedy, music, sport, television and film.

### How you need to respond.

***Do you recognize this face?***

For each face, say YES if you recognize the face.

Recognition memory is associated with two different kinds of awareness:

- **Remember:** Quite often recognition brings back to mind something you recollect about what it is that you recognize. If you can mentally re-experience a previous occasion when you encountered the face, remembering specifically when or where you have encountered it before, then please say REMEMBER'.

- **Know:** At other times recognition brings nothing back to mind about what it is you recognize about the face. You are confident that you recognize the face, and you *know* you recognize them because of strong feelings of familiarity, but you do not recollect anything you experienced when you saw the face. In this case, please say 'KNOW'.

---

[*] this was replaced with voices for the voice condition.

For each face that you recognize, after you have given the 'YES' response, please then give the REMEMBER response if recognition is accompanied by some recollective experience, or give the KNOW, if recognition is accompanied by strong feelings of familiarity in the absence of any recollective experience.

There will also be times when you do not remember the face, nor does it seem familiar, but you might want to guess that the face belongs to a person you have encountered previously. Feel free to do this, but if your YES response is really just a guess, please then give the 'GUESS' response.

Each face will be presented to you for approximately 7 seconds. You will encounter each face once throughout the experiment and you can not change your decision once you have made it. You will be give 20 seconds to make your decision.

# APPENDIX 5.

Descriptions of participants' responses from Experiment 4a (Chapter 6).

| [*]Participant No. | Remember | Know | Guess |
|---|---|---|---|
| 1 | *Chris Tarrant:* Millionaire programme. Recently being interviewed about the Major scandal. | *Bruce Forsyth:* TV show. | *Sally Gunnell:* Sport maybe. Don't know why. |
| | *Johnny Vegas:* Shooting stars and Digital TV. | *Noddy Holder:* Definitely recognize. Maybe the singer from Slade, but looks very old in this clip. | *Chris Eubank:* Boxer. Not sure, maybe because I saw somebody that looked like him earlier. |
| 3 | *Linford Christie:* 100m sprinter. | *Michael Portillo:* Associated with politics. Definitely politician, maybe conservative party. | *Ricky Tomlinson:* Not sure. |

---

[*] = Participant numbers 1, 3 and 14 are from the standard face condition. Numbers 7, 35 and 52 are from the blurred face condition and numbers 8, 12 and 13 are from the voice condition.

| | | | |
|---|---|---|---|
| | *David Beckham:* Manchester United and England player. Always in press. | | *Ann Widdecomb:* Politician. Guess. |
| 14 | *Dawn French:* Own clothes range. Funny actress and married to Lenny Henry. | *Kirsty Young:* On TV | *Sophie Ellis Bextor:* Very pretty and I assumed that I must have seen her in the press. Maybe a model. |
| | *Eammon Holmes:* Chat show host. Morning TV. | *Ronan Keating:* Irish Boyband. Don't know which one because they all look alike. | |

| | | | |
|---|---|---|---|
| 7 | *Michael Barrymore:* In the newspapers recently about the murder case. Dead man found in swimming pool. | *Ricky Tomlinson:* Somebody from TV. | *David Beckham:* Don't know. |
| | *Kate Winslet:* Actress in Titanic. | | |
| 35 | *Linford Christie:* Sprinter. Involved in drugs test scandal. | *Liam Neeson:* Very familiar. I know that I know him, but I just don't know who he is. | NO GUESSES |
| | *Rod Stewart:* Scottish singer. In the tabloids because he dates young, beautiful, blond women. | *Michael Heseltine:* The first time I saw him, I knew I recognized him and I associated him with John Major, but its not John Major. | |

| | | | |
|---|---|---|---|
| 52 | *Ronnie Barker:* Comedian. Double-act. May have passed away recently. | *Ronnie Wood:* Musician. Don't know anything else. | *Julie Walters:* Don't know |
| | *Anne Robinson:* Weakest Link. | *Vinnie Jones:* Thought I recognized him from a film. He just reminded me of 'hard man' roles. Then Lock, Stock and two Smoking barrels came to mind. | *David Beckham:* I thought of football. Maybe David Beckham, but I wasn't sure because he always changes his hairstyle and he looks very different know. |
| 8 | *Anne Robinson:* Watchdog and Weakest Link. | *Vinnie Jones:* Really familiar voice. Very strong Eastend accent and I thought it could either be Vinnie Jones or Shane Ritchie. Definitely a familiar voice. | *Chris Eubank:* Don't know. |

| | | | |
|---|---|---|---|
| | *Bruce Forsyth:* Generation Game. | *Ann Widdecomb:* Politics. Similar to Mo Mowlam, but I know that it is not her. I can't picture her face. | *John Major:* Don't know. |
| 12 | *Terry Wogan:* TV presenter who does charity work. Pudsy Bear. | *Sally Gunnell:* Don't know. | *Margaret Thatcher:* Old lady with posh voice. I thought somebody with such a posh voice must be important and famous for something. |
| | *Bob Monkhouse:* Presents gameshow on BBC 1 after Neighbours. It is not Des O'connor. | *Carol Voderman:* Definitely on TV. | *John Major:* Again it was the posh accent. I guessed that I recognized him. Maybe he's a stage actor. |
| 13 | *Ann Widdecomb:* Politics. In the media recently for having a make-over. She is not very pretty. | *Lulu:* Don't know. | *Jarvis Cocker:* Don't know. |

| | | |
|---|---|---|
| *Dawn French:* French and Saunders. Big lady. Vicar of Dibley. | *Boy George:* Don't know. | *Esther Rantzen:* Guess. |