Article:

Davies, Graham and Young, Andrew William orcid.org/0000-0002-1202-6297 (2017) Research on face recognition:the Aberdeen influence. British journal of psychology. pp. 812-830. ISSN 0007-1269

https://doi.org/10.1111/bjop.12243

Authors' version as accepted for publication in *British Journal of Psychology*, January 19th 2017.

Research on face recognition: the Aberdeen influence

Graham M. Davies and Andrew W. Young

Department of Psychology, University of Leicester, UK

Department of Psychology, University of York, UK

Authors' Note

This paper has its origins in a very successful symposium on 'Face Processing: Past, Present and Future' held at the BPS Cognitive Psychology Section Annual Conference at the University of Kent in September 2015. The symposium was convened by Bob Johnston and the late Allan McNeill to bring together some of the old-timers and up and coming younger researchers to take an overview of this thriving field of investigation. The speakers included Graham Davies, Lisa DeBruine, Stefan Schweinberger, Harriet Smith, Vicki Bruce, Tim Valentine, Sarah Stevenage, Rick Hanley, Andy Young, David Etchells, Karen Lander, Mike Burton, Lubna Ahmed, David Perrett, Hannah Tummon, Allan McNeill, Markus Bindemann, and keynote Jim Tanaka. We know that Bob and Allan were keen that at least some of the things discussed at Kent should find their way into the *British Journal of Psychology*, which has for many years been a leading vehicle for disseminating high quality research on face recognition, and we drafted this paper with this background very much in mind.

We are grateful to Stefan Schweinberger, who kindly looked at our first attempt and encouraged us to be ambitious in our goals. We also thank Holger Wiese, David White and an anonymous reviewer for their helpful comments on a previous draft. The authors take equal responsibility for the content of this paper; order of authorship was determined by age.

Correspondence to Graham Davies, School of Neuroscience, Psychology and Behaviour, Centre for Medicine, University of Leicester, University Road, Leicester LE1 7RH. E-mail:gmd@le.ac.uk

Research on face recognition: the Aberdeen influence

Abstract

The review of 'Recognizing faces' by Hadyn Ellis, published in the *British Journal of Psychology* in 1975, marked the genesis of a distinct field of research. This seminal review sprang from a broader programme of research on face recognition conducted at the University of Aberdeen, whose influence continues to be felt in what has become an internationally important research area. We discuss the background to the Aberdeen research, summarise some of its achievements, and offer reasons why it proved so successful. These reasons include the synergy between theory and practice-based studies, the value of combining different perspectives and sources of evidence, sound techniques that led to easily reproducible findings, and the emphasis on testing even the most 'common sense' assumptions.

Research on face recognition: the Aberdeen influence

Face recognition and face perception are now mainstream topics in psychological research, but it wasn't always that way. In the early 1980s, Goldstein (1983) had pointed out that a remarkable array of behavioural scientists had published one or two studies on faces and then moved on to other topics, leaving a legacy of disconnected work that didn't form a coherent whole. A paper that played a key role in changing that state of affairs was an influential review by Hadyn Ellis of research on '*Recognizing faces*', published in the *British Journal of Psychology* (Ellis, 1975). In retrospect, this review marked the genesis of a distinct field of research which is acknowledged today as internationally important and continues to represent an area of strength in British psychological science (Calder et al., 2011; Schweinberger & Burton, 2011; Bruce & Young, 2012; Young & Burton, in press).

Ellis's (1975) review paper was part of a very successful programme of research on faces at the University of Aberdeen. Some forty years later it is timely, perhaps, to look back at the beginnings of the research instigated and led by Hadyn Ellis and his colleagues, and the influence it still has on today's studies of face perception and indeed more widely. In looking back, we aim also to show some of the factors that contributed to the success of the research. These include the synergy between theory and practice, the value of combining different perspectives and sources of evidence, sound techniques that led to easily reproducible findings, and the emphasis on testing even the most 'common sense' assumptions. Because much of the work was built around investigating the then newly introduced Photofit kit, we have used this to frame our discussion.

Recognizing Faces (Ellis, 1975)

Ellis's (1975) review brought together previously disparate lines of evidence concerning face recognition. It included a comprehensive analysis of what was then known about stimulus and participant factors which affect face recognition, and it considered evidence from experimental research on face memory, developmental studies of infants and children, and clinical studies of inability to recognise faces (prosopagnosia).

A number of the topics discussed in the review have entered the mainstream. These included the possibility of Gestalt (nowadays usually called holistic) perception (Young et al., 1987; Tanaka & Farah, 1993; Maurer, LeGrand & Mondloch, 2002; Rossion, 2013), the dominance of brightness over colour information in face recognition (Kemp et al., 1996), the salience of features in the eye region (Caldara et al., 2005; Gilad, Meng & Sinha, 2009; Peterson & Eckstein, 2012), the other-race effect and the possibility that different cues are optimal for recognising faces of different ethnicities (Rossion & Michel, 2011; Bruce & Young, 2012), and what can be learnt from the impact of unusual transformations such as inversion of an image's orientation or its contrast relationships (Kemp et al., 1996; Leder & Bruce, 2000; Gilad et al., 2009; Kramer et al., in press).

Finding and synthesising such a diverse literature was in itself a considerable achievement in the days before internet searches would make the task so much easier, but Ellis (1975) also identified themes that became the focus of many later studies. With hindsight, three stand out particularly clearly; the question of whether or not faces have a 'special' status, the importance of neuropsychological evidence, and the role of theory.

In discussing whether or not faces are 'special' Ellis (1975) was careful to distinguish the different senses in which faces might be considered special (later elaborated by Ellis & Young, 1989) and show how different lines of evidence were relevant to each type of claim. His overall conclusion was that faces unquestionably have a special psychological status in our everyday lives, but that evidence at the time was inadequate to conclude that the processes involved in recognising faces are qualitatively different from those used in recognising other pictorial material. Ellis's (1975) discussion of these issues was both more cautious and more sophisticated than some that have followed in what has become quite a polarised debate.

Neuropsychological studies discussed by Ellis (1975) included the effects of unilateral cerebral lesions and the severe deficit of familiar face recognition found in cases of prosopagnosia. At the time, this use of neuropsychological alongside experimental evidence was unusual, though the work of Yin (1969, 1970) in Hans-Leukas Teuber's lab at MIT was a notable exception. The case reports of prosopagnosia were in journals most psychologists did not consult, and the patients were rare. Ellis (1975) pointed out that only Henri Hécaen’s group in France had studied a number of patients, but these were tested at different time periods with varying procedures. Moreover, structural neuroimaging methods were limited to non-existent. Without good neuroimaging evidence, lesion sites often had to be inferred from indirect evidence, and Ellis (1975) noted case reports where prosopagnosia had been thought to follow damage to very variable brain regions. He commented that this diversity of putative lesion sites might suggest that face recognition involves ‘a number of scattered sub-processes any of which, if interfered with, can lead to a total inability to recognize faces’ (Ellis, 1975, p. 421). The sceptic’s alternative, though, was that much of the localisation information was imprecise or simply incorrect. In fact, a remarkably astute contemporary review of the neuroanatomical basis of prosopagnosia by Meadows (1974) offered a different emphasis, identifying lesions in the right inferior occipito-temporal junction as critical, largely from a careful analysis of visual field defects in reported cases and the small number of post-mortem reports. It turned out that Meadows' (1974) neuroanatomical analysis was largely correct (Young, 2011), but Ellis (1975) none the less deserves credit for seeing the importance of a functional (rather than purely neuroanatomical) analysis of the problem in recognising faces. In making this link, Ellis (1975) argued that both normal and abnormal face memory must reflect the same underlying cognitive mechanisms. This became one of the cornerstones of the field, and indeed of cognitive neuropsychology more generally (Hay & Young, 1982; A. Ellis & Young, 1988; Young, 2011).

On the role of theory, Ellis (1975, p.417) commented that 'One very striking aspect of the research on face recognition reviewed so far is the absence of any unifying theoretical structures'. Given that few people at the time saw face recognition as a distinct field of enquiry, this was perhaps unsurprising, and various attempts have since been made to create more coherent perspectives (Hay & Young, 1982; Bruce & Young, 1986, Valentine, 1991; Burton, Bruce & Hancock, 1999). Ellis (1975) suggested that accounts of perceptual learning might help fill this gap, and especially Eleanor Gibson's (1969) concept of perceptual differentiation. This suggestion was partly sound and (as it turned out) partly misleading.

The sound part was the emphasis on the potential value (still not fully realised) of an ontogenetic perspective (Johnson, 2011; Karmiloff-Smith, 2013). The part that turned out to be misleading was the emphasis on perceptual differentiation. This followed directly from the assumption (which can be traced back to Galton, 1883) that faces form a particularly homogeneous class of visual stimuli, with reduced susceptibility to verbal coding (Ellis, 1975, p.409). From this assumption, nearly everyone inferred that the key to successful face recognition is the ability to visually discriminate between different faces. That is, the core problem was thought to be one of using relatively subtle differences to tell different faces apart. Many researchers still hold this position, and of course discrimination between different individuals is indeed a necessary part of face recognition, but important studies by Mike Burton and his colleagues have clearly exposed the limitations of an exclusive focus on this property (Burton, 2013).

To understand Burton's work we need to consider a weakness of widely used experimental paradigms that conflate picture recognition and face recognition by testing the recognition of photographs of faces that are identical to the photographs that were previously studied. Nearly all the studies of recognition memory for faces reviewed by Ellis (1975) were of this type. Such tasks led to high levels of performance (typically over 90% correct) that contrasted markedly with an increasing awareness of witness misidentifications from everyday life (Shepherd, Ellis & Davies, 1982; Davies & Griffith, 2008). The apparent paradox was resolved by Vicki Bruce's (1982) demonstration that laboratory performance falls off quickly when any change is introduced between studied and test face photos; it turns out that whereas recognition of particular photographs (i.e.

recognition of the same photos presented at study and test) is remarkably good, memory for unfamiliar faces (when measured across different study and test items) is comparatively poor and highly image-dependent (Longmore, Liu & Young, 2008).

This failure to distinguish picture recognition from face recognition also contributed to a widespread tendency to treat differences between images of the same face as irrelevant 'noise', leading to an emphasis on studies using highly standardised sets of images that removed as far as possible many of the sources of variability in the images of faces we actually encounter in our everyday lives (differences in lighting, pose, expression, and so on).

One of the first people to see that the limitations of this approach was Vicki Bruce (1994), who suggested instead that image variability might actually play an important role in developing stable representations of familiar faces. This theme has been developed in a number of recent studies by Mike Burton and his colleagues (Jenkins, White, van Montfort & Burton, 2011; Burton, 2013; Andrews et al., 2015). For example, Jenkins et al. (2011) introduced a simple sorting task in which participants were shown 40 everyday photographs of unfamiliar faces and asked to sort these into piles containing the same identity. Although the photographs in the set were derived from only two face identities, participants often created as many as nine different piles; that is, they mistook the differences between different images of the same face for differences in face identity. In contrast, they rarely sorted images of the two different faces into the same pile. These findings show how easily the variability between different images of an unfamiliar face can be misinterpreted as due to differences in identity. Put more generally, this problem isn't adequately characterised as one of discriminating between faces. Instead, it is as much a problem with telling different views of faces together as with telling the faces apart.

Although Ellis (1975) did not make these key theoretical points about image variability and the differences between picture recognition and face recognition, he came close. In particular, he was sceptical of findings with unrealistic 'identikit' stimuli and suggested that 'A more representative paradigm for establishing the everyday processing of faces would be one in which target faces (of well-known public figures perhaps) occur within an array of other faces and subjects are required to search through looking for any one of the targets' (Ellis, 1975, p.418). In making this point he also touched on the key issue of face familiarity. The theoretical importance of differences between familiar and unfamiliar faces was underscored in Ellis, Shepherd and Davies' (1979) study of the roles of internal and external features and in the work of Bruce (1979), who followed up Ellis's (1975) advocacy of the visual search paradigm. In Jenkins et al.'s (2011) previously described sorting task, participants seldom make any errors if the faces are familiar, and understanding differences between familiar and unfamiliar faces has become a cornerstone of much theoretical development (Hay & Young, 1982; Bruce & Young, 1986; Johnston & Edmonds, 2009; Burton, Jenkins & Schweinberger, 2011; Young & Burton, in press). The differential salience of internal features of familiar faces first noted by Ellis et al. (1979) has remained one of the key observations to be explained (Young et al., 1985b), and can even be used as a metric for face familiarity (Clutterbuck and Johnston, 2005).

Origins of the Aberdeen Research

Part of the reason why Ellis (1975) was so successful in synthesising and discussing such a diverse range of studies of face recognition was that there was already a strong background of relevant work at Aberdeen. The first published studies on faces from Aberdeen were a collaboration between the cognitive psychologist, Hadyn Ellis and the social psychologist, John Shepherd. They actually began by studying facial attractiveness (Shepherd and Ellis, 1973) - a topic which has itself became one of the mainstream issues in face perception - but Hadyn always credited John with the insight that the deceptively simple question as to how we recognise faces might be worth pursuing in its own right.

Subsequently, they were joined by Graham Davies who had interests in memory for the structure and content of pictures. The combined expertise of this core team was put to good use when the issue of the usefulness of '*Photofit*' was raised: this was a new tool adopted by the British police to assist witnesses in recalling the facial appearance of offenders. *Photofit* had received a good deal of media publicity and was in widespread use, but there was no actual research to demonstrate its effectiveness. The team decided to explore this issue, which became a focus for much of the Aberdeen work and its main source of external funding.

Early Trials with *Photofit*

Rather to the researchers' surprise, the Police Scientific Development Branch (PSDB) agreed to lend the Aberdeen team a male *Photofit* kit. It consisted of a set of around 600 photographic representations of facial features, which together with some accessories (hats, eyeglasses etc.) could be mounted onto a specially designed frame to create a composite face. The majority of the kit's parts showed varieties of five key features: hairlines and ears, eyes and eyebrows, noses, mouths, and chins. With the assistance of an operator, witnesses selected their preferred features of the remembered face from the 'Visual Index': an album containing miniature reproductions of all of the components in the kit. The resulting composite could then be improved by exchanging features until an image acceptable to the witness emerged (Graham Pike has produced an excellent working demonstration of *Photofit* and the quality of likeness it can achieve, which can be viewed at http://www.open.edu/openlearn/body-mind/photofit-me).

The police had begun to adopt Photofit because its validity seemed self-evident: it fitted the common sense view that the parts of faces comprise eyes, noses, mouths and so on. However, putting aside this intuitive approach, the Aberdeen team conducted some basic evaluation studies which instead suggested, somewhat counterintuitively, that the existing kit might have rather limited potential to produce an identifiable likeness of a suspect (Ellis, Shepherd & Davies, 1975). Volunteers attempted to make a copy of a white male face which was itself made from *Photofit* features, either from brief exposure to the composite 'face' or with the 'face' continuously present for reference. Participants found these tasks rather difficult; remarkably, even with the composite *Photofit* target 'face' in full view, no-one succeeded in completing it correctly.

In a second experiment, the same participants attempted to make *Photofit* likenesses of real faces, each seen briefly as photographs. No two participants selected the same features for any face and the degree of resemblance achieved varied widely (see Figure 1). Independent judges, who sought to match each composite to the actual face in a mugshot array, were accurate just 1 in 8 times; this ratio improved to 1 in 4 when second and third choices were taken into account. Those participants who had produced more accurate copies of the *Photofit* face in the first experiment produced significantly better likenesses of real faces in the second task: the only individual difference variable that the team ever found which predicted composite production accuracy. Overall, these results suggested that under experimental conditions, the majority of Photofits bore, at best, only a limited resemblance to their target.

--Figure 1 about here---

When results from these first studies appeared in the *British Journal of Psychology* (Ellis, Shepherd & Davies, 1975), the story was picked up by *The Times.* Its unfortunate headline: "Photofit 'useless' say scientists" led to threats of legal action from the system's inventor, Jacques Penry, and some embarrassment in PSDB and the Home Office. It was against this background that Ellis (1975) published his seminal review of existing research on face recognition.

Later *Photofit* Studies

In 1974, the Aberdeen team successfully sought funding from the then Social Science Research Council (now, the ESRC) to explore further the potential of *Photofit*. These later studies confirmed the insensitivity of the kit to factors known to influence memory performance in general, such as exposure time and delay. Most strikingly, no consistent differences emerged in assessed quality between composites made in the presence of a target compared to those made from memory. This was taken at the time to reflect limitations of the kit itself. That interpretation was no doubt largely correct, but with hindsight the findings also pointed to now well-established problems in unfamiliar face perception (Hancock, Bruce & Burton, 2001; Megreya & Burton, 2006). As already noted, our visual systems find it surprisingly difficult to determine whether differences between images of unfamiliar faces reflect differences in face identity or are instead, due to identity-irrelevant picture differences (Hancock et al., 2001). In the example we gave previously of Jenkins et al.'s (2011) photograph sorting task there is no memory component; all the photographs are visible throughout and can be looked at for as long as the participant requires. Even so, substantial errors are made if the faces are unfamiliar. Hence a problem we often think of as one of face memory (such as errors in eye witness testimony with unfamiliar faces) can actually be as much one of *perceiving* images of unfamiliar faces. The otherwise puzzling finding that volunteers in the Ellis et al (1975) study could not even copy a Photofit 'face' correctly is also consistent with this view. Again, the importance of testing even the most basic assumptions is clear; in this case the assumption that perception of the identities of unfamiliar faces is unproblematic.

It took some time for researchers to recognise that there might be substantial problems in unfamiliar face perception, probably because the ease with which we can carry out such tasks with familiar faces misleads us into thinking we are excellent at recognising all faces and thus makes the finding of problems with unfamiliar faces strongly counterintuitive (Ritchie et al., 2015). Yet important field studies have confirmed that unfamiliar face matching is remarkably error-prone in contexts as varied as passport control and the supermarket check-out (Kemp, Towell & Pike, 1997; White et al., 2014) and that errors can be made even by highly-trained and very experienced staff (White et al., 2014)

The Aberdeen research identified a number of weaknesses in the *Photofit* kit*,* which it shared with rival composite systems like the US *Identikit*: these included gaps and duplications in its range of features and problems in locating relevant features from witnesses' descriptions. Surprisingly, the initial verbal descriptions which formed the basis of the *Photofit* composites provided a better guide to likeness than the composites themselves (Christie & Ellis, 1981), underlining the importance of careful interviewing to elicit the maximum amount of useful information. Further research showed that a skilled operator spent considerably longer establishing a verbal description than did a novice operator, before turning to the kit itself (Davies, Milne & Shepherd, 1983). Today's composite systems also lay emphasis on eliciting the initial witness description, through such mnemonic techniques as the Cognitive Interview (Frowd, Bruce, Smith, & Hancock, 2008). More fundamentally the Aberdeen studies brought into question the view that witnesses necessarily remembered faces in a way which enabled them to assemble a convincing composite face from its component features (Davies, 1981; Shepherd & Ellis, 1996). This suggested the need for a different approach to constructing facial composite systems and dovetailed with later research demonstrating the importance of holistic perception of faces (Young, Hellawell & Hay, 1987; Tanaka & Farah, 1993; Rossion, 2013), a point we return to later.

Initial embarrassment at PSDB soon gave way to constructive engagement: it emerged that the police, too, had concerns about the effectiveness of *Photofit*, with reports of failures and weaknesses paralleling those uncovered in the Aberdeen research. Their reaction to such findings was admirably positive: from 1975 to 1987, PSDB (subsequently SRDB: the Scientific Research and Development Branch, Home Office) provided funding to the Aberdeen group not only to improve the quality of the existing *Photofit* technology, but also to undertake basic research on how faces were perceived and recognised to enable future composite systems to start from a sound psychological analysis.

Practice-Based Research

The Aberdeen *Photofit* studies described above were mainly targeted at fundamental questions about the adequacy of the system, but the group's efforts also turned to how its use might be improved in practical contexts. Research designed to maximise the potential of the existing *Photofit* system covered such issues as the mechanics of image building. As noted earlier, the starting point for creating the *Photofit* composite images required witnesses to select features from small photographs in the 'Visual Index'. The Aberdeen team were able to demonstrate that the accuracy of judgements for features in isolation was less veridical than judgements of the same features seen in the context of a composite face (Davies & Christie, 1982); a finding that parallels the procedure later used by Tanaka and Farah (1993) to demonstrate holistic perception. In addition, the boundary lines delineating features in completed Photofit composites impaired identification, perhaps by disrupting normal scanning patterns for faces (Ellis, Davies & Shepherd, 1978). Through eliciting how people actually describe facial features and mapping their description onto the examples featured in the Visual Index, the team were able to restructure the Index in a way which eased the task of the operator in locating relevant features. The range of features themselves was also expanded and enhanced by cannibalising the companion *Penry Female Photofit* kit for some of its more androgynous features. These new features enabled operators to build more convincing young male faces, a common cause of frustration voiced by police operators using the conventional kit (the elements in the original male kit were taken from photographs in criminal records, which inevitably included a disproportionate number of seasoned offenders). The culmination of this development work was the *Aberdeen Index to Photofit* which was taken up by a number of police operators in their quest for a better quality of likeness. Jean Shepherd, who had joined the Aberdeen team initially as the *Photofit* operator and Donald Christie, the team's Research Fellow, played invaluable roles in this development work.

Complementing this laboratory research, the team collaborated with SRDB on a survey of the experience of serving police operators with *Photofit* (Kitson, Darnbrough & Shields, 1978). This survey revealed the need, not just for training covering the uses and limitations of the kit, but also broader input from psychologists on what was known about face recognition and witness memory in general (Davies, Shepherd, & Ellis, 1978). In 1977, the Aberdeen team collaborated with experienced police operators in the first of what proved to be a long-running series of annual courses for police officers dealing with establishing suspect identity from witness descriptions (Davies, Shepherd, Shepherd, Flin, & Ellis, 1986). The involvement of cognitive psychologists in specialised police training in witness interviewing continues today, for instance at the Centre for Policing Research and Learning run by the Open University and the Centre for Forensic Interviewing at Portsmouth University (see Walsh, Oxburgh, Redlich, & Myklebust, 2016 for a recent review).

Research on Face Perception

The Aberdeen team also conducted fundamental research on what attributes in a face are likely to be particularly salient for witnesses. Clearly, any successful face recall system should give priority to modelling such information. The then novel procedure of multi-dimensional scaling (MDS) was applied in an attempt to tease out the intuitive dimensions witnesses use to classify unfamiliar faces. Sets of up to 100 photographed faces were given to participants to sort into groups on the basis of perceived similarity, which yielded a matrix showing how often one face was sorted with any other. This similarity matrix was then analysed by the MINISSA multi-dimensional scaling program to produce up to six dimensions. The three-dimensional solution which provided a satisfactory fit to one data set is shown in Figure 2. As the features and general appearance of the same faces had previously been rated on a large number of scales, it was possible to label the dimensions participants had employed in making their initial judgements. The three dimensions most frequently found were for *age*, contrasting lined and balding men with smooth-faced youths; *face shape*, a dimension emphasising face length and fatness and *hair length* contrasting long-haired students (this was the 1970's!) with men with well-groomed, shorter hair (Shepherd, Davies & Ellis, 1977). These dimensions appeared to rest not just on particular features, but on overall impressions of a face, something that traditional composite systems had difficulty capturing.

---Figure 2 about here---

Although related approaches had been used to evaluate the underlying similarity structure of facial expressions (Schlosberg, 1954; Woodworth & Schlosberg, 1954), the Aberdeen work was strikingly ahead of its time in addressing more general questions of similarity in appearance. Dimensional approaches to face identity perception have since become very popular. A particularly widely used idea has been Valentine's (1991, 2001) 'face space' metaphor. According to Valentine, Lewis & Hills (2016), the concept of a multi-dimensional 'face space', in which faces can be represented according to their perceived properties is 'fundamental to the modern theorist in face processing' (p.1996). However, Valentine's (1991, 2001) concept of a space with high dimensionality has some limitations in that the dimensions remain unknown and also that multidimensional spaces have counterintuitive properties, as noted by Burton & Vokey (1998). These limitations make it difficult to draw strong predictions from the model, which tends instead to be used as an interesting analogy. In contrast, the pioneering work on MDS by the Aberdeen team identified a relatively small number of dimensions that could be quantified and fitted into a conventional geometric representation. A closer analogue may therefore lie in the way that it presaged modern dimensional approaches to understanding facial first impressions (Oosterhof & Todorov, 2008; Sutherland et al., 2013), as shown here in Figure 2. These approaches use similar techniques to identify two or three dimensions that can encapsulate a wide range of attributes that underlie our first reactions to the faces of strangers and are capable of modelling much of the variance linked to these impressions (Vernon et al., 2014).

The similarity matrices constructed at Aberdeen could also be subjected to Hierarchical Clustering Analysis to identify groups of faces which shared key similarities while differing on other, less salient dimensions. Research with face recognition tasks demonstrated that between 70 and 80% of errors made by participants could be accounted for by confusions among faces from the same clusters, offering a rational basis for mistakes in identification (Davies, Shepherd & Ellis, 1979).

The advent of the Devlin report (Devlin 1976) into miscarriages of justice offered the Aberdeen team an opportunity to study mistaken identification errors in the formal context of police identification parades. During the compilation of his report on the causes and consequences of witness error, Devlin had taken evidence from psychologists as well as lawyers. He was sufficiently impressed by the relevance of psychological research to the problems of the courts that in his report he directed that ‘the insights of psychology’ should be brought to bear ‘on the conduct of parades and the practice of the courts’ (p.149). When the Home Office announced a fund for such research, the Aberdeen team were well placed to apply successfully for financial support to investigate the impact of delay on identification accuracy: long delays between observation by witnesses of suspects and their appearance at an identification parade had been a significant feature of some of the cases reviewed by Devlin.

As the research was to be aimed at the courts, ecological validity was prioritised in the Aberdeen research: unexpected incidents were staged in front of unsuspecting members of the public; selection and conduct of identification parades were overseen by serving police officers, and the delays between initial observation by the witness and the holding of a parade were extended by intervals of up to one year, still among the longest delays in witness testimony experiments (Shepherd, Ellis & Davies, 1982). Figure 3 illustrates the outcome of one of these studies. As can be observed, the number of witnesses who made a positive and correct identification declined steadily over time, but the rates of error (picking an innocent stand-in) remained remarkably constant; by one year, the great majority of participants opted for no choice, an option introduced explicitly in the new parade instructions instituted by Devlin.

---Figure 3 about here--

Wider Horizons in Face Processing

The achievements of the research on adult face recognition also stimulated a range of other approaches at Aberdeen. For example, Andy Young and Dennis Hay followed up Ellis and Shepherd's (1975) study of cerebral lateralisation (Young & Ellis, 1976; Hay, 1981) and Jan Deregowski added a cross-cultural perspective (Shepherd, Deregowski & Ellis, 1974; Ellis, Deregowski & Shepherd, 1975). Hadyn Ellis had from the start seen the value of a developmental perspective (e.g. Ellis et al., 1973;

Young & Ellis, 1976). He supervised research by Rhona Flin and others on the development of face recognition ability and its apparent regression in early adolescence (Flin, 1985). In a particularly notable pioneering effort, Hadyn recognised the potential importance of a preliminary report of neonatal responses to face-like images by Goren, Sarty and Wu (1975). He worked with Suzanne Dziurawiec to carefully investigate these neonatal responses in studies that were later included in the highly-cited paper by Johnson, Dziurawiec, Ellis and Morton (1991). This broader developmental portfolio again offered complementary perspectives that drove significant theoretical advances (Ellis, 1992) and fed back into the work on more directly applied questions (Flin, Markham & Davies, 1988 Davies, 1993).

Technological innovation in Face Identification Systems

In the late 1970s, scientists at SRDB recognised the potential of computers to assist recall of faces, by offering virtually infinite variations in the shape, size and relative position of features and the ability to change the overall face shape at the command of the witness. SRDB commissioned a prototype computerised composite system using *Photofit* principles from the Computer-Aided Design Centre (CADC) in Cambridge. Developing a graphics package which could display a composite face and allow the smooth interchange and manipulation of features proved unexpectedly difficult, but was finally solved (Craw, Ellis & Lishman, 1986). When the system was up and running, the Aberdeen team were asked to conduct a formal comparison of the image quality achieved by the new system in comparison with the orthodox *Photofit* kit in the hands of an experienced operator. While both systems produced more identifiable composites when made by witnesses from view than from memory of a face, there was no overall difference in the accuracy achieved by the two systems (Christie, Davies, Shepherd & Ellis, 1981). However, the potential of computers for composite construction had been demonstrated; further progress had to await the arrival of the desk-top computer and cheap, versatile graphics packages.

When the new systems arrived, some systems, such as FACES and Mac-A Mug, simply emulated the feature-based construction process of Photofit and the Identikit. Laboratory research with these new devices proved disappointing: despite the ability of such systems to manipulate and edit features within a face being considerably greater than the previous systems, the degree of likeness achieved by volunteers working from memory remained obstinately low (see Davies & Valentine, 2007; Frowd , 2015, for reviews).

The Aberdeen research had pointed to the importance of an holistic approach: of judging features in the context of the total face rather than in isolation (Davies & Christie 1982; Tanaka & Farah, 1993). This principle had been embodied in the prototype CADC system and its direct descendants, E-fit and PRO-fit: such systems begin from a schematic face generated on the basis of the witness's initial description, which is then amended by exchanging or resizing features. Davies, Van der Willik & Morrison (2000) directly compared the quality of likeness achievable with E-fit and the old Photofit: when witnesses were able to refer to photographs of the target face, the degree of likeness achieved by E-fit was altogether superior to that achieved with Photofit. However, when other witnesses worked from memory on the same faces, there was no difference in the assessed quality of likeness between the old and new technologies. It appeared that a wholly feature-based approach could not sufficiently capture holistic information stored about the face and that different approaches might be required. One clue as to what could be achieved by this different approach emerged from the last major Home Office project undertaken at Aberdeen.

Many offenders observed by witnesses will have previous convictions and their personal details will have been stored in police records. However, asking witnesses to search through mugshot albums can be laborious and time-consuming, prone to both false positives and false negatives (Davies, Ellis & Shepherd, 1979). The development of high capacity video discs in the early 1980s opened the way to more speedy presentation of facial images, but the problem of sheer number**s** was still a practical issue. Working with SRDB, the Aberdeen team developed the FACES system to address this latter problem. The database of the prototype system consisted of 1000 colour photographs of male faces, which had been coded on 48 facial dimensions using Likert scales. The initial batch was laboriously coded by hand, but a semi-automatic system was devised by John Shepherd based on facial measurements taken directly by computer from the face and then regressed onto the Likert scales. In the working system, volunteers described a face taken from the data-base to the operator who then converted it into points on relevant scales. An algorithm then searched the data base for the closest fit to the description. In trials, FACES located the target face in its 'top ten' on 80% of occasions (Shepherd and Ellis, 1996). An experimental comparison of the system with a manual search of mug-books showed the superiority of FACES, particularly on typical faces or when the target face appeared late in the mug-book searches (Ellis, Shepherd, Flin, Shepherd & Davies 1989).

A new generation of composite systems has sought to tap into and exploit the holistic information available to witnesses, using the approach based on overall initial impressions first pioneered at Aberdeen. Examples include the *EvoFIT* system developed by Charlie Frowd, Peter Hancock and others at Stirling University (see Frowd, 2015 for a review) and the *EFIT-V* Facial Recognition Software created by Soloman, Gibson and colleagues at Kent (Solomon, Gibson & Mist, 2013). Both eschew the selection of features and instead, use an evolutionary approach: witnesses look through a series of randomly-generated faces and select those whose overall appearance bears the most resemblance to the target face. A computer algorithm then randomly generates a further set of faces based on these choices from which the witness again selects the best resemblance. The process is repeated until an image emerges which the witness judges to be an acceptable likeness to the target. Such an approach eliminates the need for witnesses to describe individual features and instead seeks to capitalise on recognition rather than verbal recall. *EvoFIT* reverses the traditional emphasis in composite systems by initially focussing on internal facial features, with consideration of hair being left until the last stage. *EFIT-V* boasts controls which permit age acceleration or weight gain in a generated face. Although to our knowledge, no formal operational comparisons have been conducted on the quality of likeness achieved relative to the traditional feature-based approach, police forces which have adopted such technology report a jump in successful identifications with this new approach (see for instance https://www.kent.ac.uk/physical-sciences/Impact/cjs-impact/ref2014.html).

The Lasting Legacy

The FACES project was the last major piece of research conducted by the Aberdeen team. In 1986, Hadyn Ellis left Aberdeen to take up a Chair in psychology at UWIST and Graham Davies followed in 1987 to lead Psychology at the former North East London Polytechnic (now University of East London) before taking a Chair at Leicester University. Jean and John Shepherd continued their work at Aberdeen, helping to develop the feature library of *E-fit*. Shortly after UWIST merged with the University of Wales, Cardiff, Hadyn Ellis was appointed Head of the new Department of Psychology, just the next step in a stellar career, which eventually led to him being appointed as the University's Deputy Vice-Chancellor, with special responsibility for research. In between the demands of these onerous posts, Hadyn continued his work on face recognition at Cardiff, including his interest in delusional misidentification (e.g. Ellis & Young, 1990; Ellis et al., 1997; Bell, Halligan & Ellis, 2006) and priming effects (Ellis, Ellis & Hosie, 1993; Ellis. Jones & Mosdell, 1997). The studies of delusional misidentification in particular achieved significant influence on theories of face perception (Ellis & Lewis, 2001) and have been recognised as fundamental to cognitive neuropsychiatry, a new field whose name Hadyn had himself coined (Coltheart, 2006). His untimely death in 2006 robbed the field of one its most creative and influential researchers. The Hadyn Ellis Building of the School of Medicine at Cardiff University stands in his honour.

By the time Hadyn Ellis left Aberdeen, a number of groups devoted to face research had sprung up in other universities in the UK. Some reflected an Aberdeen diaspora – Dennis Hay and Andy Young had taken jobs at Lancaster, where they found a new colleague called Ellis (Andy Ellis) with very creative ideas (e.g. Young, Hay & A. Ellis, 1985a) - and other groups arose *de novo*. A shared recognition of common purpose in establishing this new field led to a series of workshops devoted to facial research which took place annually at Grange-over–Sands, Cumbria, where ideas were exchanged and new research and theories received an airing. The fact that the field was new probably contributed to the informal and supportive atmosphere of these meetings, and it was at Grange that the influential Bruce and Young model of face recognition (Bruce & Young, 1986) was developed as a collaborative effort by several individuals (see Schweinberger & Burton, 2011; Young & Bruce, 2011).

In the years when they worked together, the Aberdeen team published over 100 scientific papers and four books on face processing. In our opinion, a variety of factors contributed to the success of the research. Central was the link between theory and practice; while much of the government funding given to the Aberdeen group was for applied issues, basic research on face processing was never neglected and often went hand in hand. In many ways, they established what have become essential characteristics of the best psychology research, including the impact of a multifaceted approach combining different perspectives and sources of evidence. Moreover, they showed the value of directly addressing assumptions that can seem so intuitively obvious that they often get overlooked. For example, no-one had thought to question whether a competent observer would be able to recreate a plausible likeness of a seen face from the components of the *Photofit* kit. In particular, though, the Aberdeen work demonstrated compellingly the synergy between theoretical and practical questions, in which theory informs practice but practice forces the development of better theories in equal measure. The underlying technical qualities of the Aberdeen studies were sound, too. In the modern era, where the poor reproducibility of some findings in psychology has created much discussion (Open Science Collaboration, 2015), the ease of replicating key findings from the Aberdeen work stands out. For such reasons, its influence still resonates in what is now an international field of research.

References

Andrews, S., Jenkins, R., Cursiter, C., & Burton, A. M. (2015). Telling faces together: learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology, 68*, 2041-2050. doi: 10.1080/17470218.2014.1003949

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). Explaining delusions: a cognitive perspective. *Trends in Cognitive Sciences, 10*(5), 219-226. doi: 10.1016/j.tics.2006.03.004

Bruce, V. (1979). Searching for politicians: an information processing approach to face recognition. *Quarterly Journal of Experimental Psychology, 31*, 373-395. doi: 10.1080/14640747908400734

Bruce, V. (1982). Changing faces: visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*, 105-116. doi: 10.1111/j.2044-8295.1982.tb01795.x

Bruce, V. (1994). Stability from variation: The case of face recognition. *Quarterly Journal of Experimental Psychology, 47*, 5-28. doi: 10.1080/14640749408401141

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327. doi: 10.1111/j.2044-8295.1986.tb02199.x

Bruce, V., & Young, A. (2012). *Face perception*. Hove: Psychology Press.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*, 1467-1485. doi: 10.1080/17470218.2013.800125

Burton, A. M., & Vokey, J. R. (1998). The face-space typicality paradox: Understanding the face-space metaphor. *Quarterly Journal of Experimental Psychology, 51A*, 475-483. doi: 10.1080/713755768

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: a model of familiar face recognition. *Cognitive Science, 23*, 1-31. doi: 10.1207/s15516709cog2301_1

Burton, A.M., Jenkins, R. & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*, 943-958. doi: 10.1111/j.2044-8295.2011.02039.x

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience, 17*, 1-15. doi: 10.1162/089892905774597254

Calder, A. J., Rhodes, G., Johnson, M. H., & Haxby, J.V. (2011). (Eds.) *The Oxford handbook of face perception*. Oxford: Oxford University Press.

Christie, D., Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1981). Evaluating a new computer-based system of face recall*. Law and Human Behavior, 5,* 209-218. doi:10.1007/BF01044764

Christie, D. F. M., & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66,* 358-363. doi: 10.1037/0021-9010.66.3.358

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17*, 97-116. doi: 10.1080/09541440340000439

Coltheart, M. (2007). The 33rd Sir Frederic Bartlett Lecture: Cognitive neuropsychiatry and delusional belief. *Quarterly Journal of Experimental Psychology, 60*, 1041-1062. doi: 10.1080/17470210701338071

Craw, I., Ellis, H. D., & Lishman, J. (1987). Automatic extraction of face features, *Pattern Recognition Letters*, *5*, 183-187. doi: 10.1016/0167-8655(87)90039-0

Davies, G. M. (1981). Face recall systems. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 227-250). London: Academic Press.

Davies, G. M. (1993). Children's memory for other people: An integrative review. In C. Nelson (Ed.), *Memory and affect in development: The Minnesota symposium on child development* (pp.123-158). Hillsdale, NJ: Erlbaum. doi: 10.1007/BF01499028

Davies, G. M., & Christie, D. (1982). Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology, 67,* 103-109. doi: 10.1037/0021-9010.67.1.103

Davies, G. M., & Griffiths, L. (2008). Eyewitness identification and the English courts: A century of trial and error. *Psychiatry, Psychology and Law,15*, 435-449. doi: 10.1080/13218710802101605

Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1979). Effect of interpolated mugshot exposure on accuracy of eyewitness identification. *Journal of Applied Psychology, 64*, 232¬237. doi: 10.1037/0021-9010.64.2.232

Davies, G. M., Milne, A., & Shepherd, J. (1983). Searching for operator skills in face composite reproduction. *Journal of Police Science and Administration, 11*, 405-409.

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1978). Remembering faces: Acknowledging our limitations. *Journal of the Forensic Science Society, 18,* 19-24. doi: 10.1016/S0015-7368(78)71178-3

Davies, G. M., Shepherd, J. W., & Ellis, H. D. (1979). Similarity effects in face recognition. *American Journal of Psychology, 92*, 507-523. doi: 10.2307/1421569

Davies, G. M., Shepherd, J. W., Shepherd, J., Flin R., & Ellis, H. D. (1986). Training skills in police Photofit operators. *Policing, 2,* 35-46.

Davies, G. M., & Valentine, T. (2007). Facial Composites: Forensic utility and psychological research. In R. C. L Lindsay, D. F Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of witness psychology:*

Volume 2: Memory for people (pp.59-86). New Jersey: Erlbaum.

Davies, G. M., van der Willik, P., & Morrison, L. (2000). Facial composite production: A comparison of mechanical and computer-driven systems. *Journal of Applied Psychology, 85*, 119-124. doi:10.1037/0021-9010.85.1.119

Devlin, P. (1976). *Report to the Secretary of State for the Home Department of the Departmental Committee on evidence of Identification in criminal cases*. London: HMSO.

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology.* Hove: Psychology Press.

Ellis, H. D. (1975). Recognizing faces: A review. *British Journal of Psychology, 66*, 409-424. doi: 10.1111/j.2044-8295.1975.tb01477.x

Ellis, H. D. (1992). The development of face processing skills. *Philosophical Tranactions of the Royal Society B, 335*, 105-111. doi: 10.1098/rstb.1992.0013

Ellis, H. D., Davies, G.M., & Shepherd, J. W. (1978). Remembering pictures of real and 'unreal' faces: Some practical and theoretical considerations. *British Journal of Psychology, 69*, 467–474. doi: 10.1111/j.2044-8295.1978.tb02123.x

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Description of white and black faces by white and black subjects. *International Journal of Psychology, 10*, 119-123. doi: 10.1080/00207597508247325

Ellis, H. D. Ellis, D. M., & Hosie, J. A. (1993) Priming effects in children's face recognition. *British Journal of Psychology, 84*, 101-110. doi: 10.1111/j.2044-8295.1993.tb02465.x

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88*, 143–156. doi: 10.1111/j.2044-8295.1997.tb02625.x

Ellis, H. D., & Lewis, M. B. (2001). Capgras delusion: a window on face recognition. *Trends in*

Cognitive Sciences, 5, 149-156. doi: 10.1016/S1364-6613(00)01620-X

Ellis, H. D., & Shepherd, J. W. (1975). Recognition of upright and inverted faces presented in the left and right visual fields. *Cortex, 11*, 3-7. doi: 10.1016/S0010-9452(75)80014-1

Ellis, H., Shepherd, J., & Bruce, A. (1973). The effects of age and sex upon adolescents' recognition of faces. *Journal of Genetic Psychology, 123*, 173-174. doi: 10.1080/00221325.1973.10533202

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1975). An investigation of the use of the Photofit technique for recalling faces. *British Journal of Psychology , 66*, 29-37. doi: 10.1111/j.2044-8295.1975.tb01437.x

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features. Some implications for theories of face recognition. *Perception, 8*, 431-439. doi: 10.1068/p080431

Ellis, H. D., Shepherd, J. W., Shepherd, J., Flin, R., & Davies, G. (1989). Identification from a computer-driven retrieval system compared with a traditional mugshot album search: A new tool for police investigations. *Ergonomics, 32,* 167-177. doi: 10.1080/00140138908966077

Ellis, H. D., & Young, A. W. (1989). Are faces special? In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 1-26). Amsterdam: North Holland.

Ellis, H.D., & Young, A.W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry, 157*, 239-248. doi: 10.1192/bjp.157.2.239

Ellis, H. D., Young, A. W., Quayle, A. H., & de Pauw, K. W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society: Biological Sciences, B264*, 1085-1092. doi: 10.1098/rspb.1997.0150

Flin, R. H. (1985). Development of face recognition: an encoding switch? *British Journal of Psychology, 76*, 123-134. doi: 10.1111/j.2044-8295.1985.tb01936.x

Flin, R., Markham, R. & Davies, G. M. (1988). Making faces: Developmental trends in the construction and recognition of face composites. *Journal of Applied Developmental Psychology, 10*, 131-145. doi: 10.1016/0193-3973(89)90001-4

Frowd, C. D. (2015). Facial composites and techniques to improve image recognizability. In T. Valentine & J.P. Davies (Eds.), *Forensic facial identification.: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 43-70). Chichester: Wiley-Blackwell.

Frowd, C.D., Bruce, V., Smith, A.J., & Hancock, P.J.B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied, 14*, 276-87. doi: 10.1037/1076-898X.14.3.276

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Appleton-Century-Crofts.

Gilad, S., Meng, M., & Sinha, P. (2009). Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences, USA, 106*, 5353-5358. doi: 10.1073/pnas.0812396106

Goldstein, A. G. (1983). Behavioral scientists' fascination with faces. *Journal of Nonverbal Behavior, 7*, 223-255. doi: doi:10.1007/BF00986268

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern-discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*, 544-549.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2001). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*, 330-337. doi: 10.1016/S1364-6613(00)01519-9

Hay, D. C. (1981). Asymmetries in face processing: evidence for a right hemisphere perceptual advantage. *Quarterly Journal of Experimental Psychology, 33A,* 267-274. doi:

10.1080/14640748108400792

Hay, D. C., & Young, A. W. (1982). The human face. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 173-202). London: Academic Press.

Johnson, M. H.. (2011). Face perception: a developmental perspective. In A.J. Calder, G. Rhodes, M.H. Johnson and J.V. Haxby (Eds.), *The Oxford handbook of face perception*. Oxford: Oxford University Press, 3-14.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*, 313–323. doi: 10.1016/j.cognition.2011.08.001

Johnson, M., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40,* 1-19. doi: 10.1016/0010-0277(91)90045-6

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: a review. *Memory, 17*, 577-596. doi: 10.1080/09658210902976969

Karmiloff-Smith, A. (2013). The 40th Sir Frederic Bartlett Lecture. Challenging the use of adult neuropsychological models for explaining neurodevelopmental disorders: developed versus developing brains. *Quarterly Journal of Experimental Psychology, 66*, 1-14. doi: 10.1080/17470218.2012.744424

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces - the role of shape from shading and pigmentation cues. *Perception, 25*, 37-52. doi: 10.1068/p250037

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: photographs, credit cards and fraud. *Applied Cognitive Psychology, 11*, 211-222. doi: 10.1002/(SICI)1099-0720(199706)11:3<211::AID-ACP430>3.0.CO;2-O

Kitson, A., Darnbrough, M., & Shields, E. (1978). Let's face it. *Police Research Bulletin*, no. 30, pp. 7-13.

Kramer, R. S. S., Jenkins, R., Young, A. W., & Burton, A. M. (in press). Natural variability is essential to learning new faces. *Visual Cognition,* doi: 10.1080/13506285.2016.1242522

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, 53A*, 513-536. doi: 10.1080/713755889

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception & Performance, 34*, 77-100. doi: 10.1037/0096-1523.34.1.77

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255-260. doi: 10.1016/S1364-6613(02)01903-4

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition, 34*, 865-876. doi: 10.3758/BF03193433

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, 6251. doi: 10.1126/science.aac4716

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, USA, 105*, 11087–11092. doi: 10.1073/pnas.0805664105

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, USA, 109*, E3314-E3323. doi: 10.1073/pnas.1214269109

Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition, 141*, 161-169. doi: 10.1016/j.cognition.2015.05.002

Rossion, B. (2013). The composite face illusion: a whole window into our understanding of holistic face perception. *Visual Cognition, 21*, 139-253. doi: 10.1080/13506285.2013.772929

Rossion, B., & Michel, C.. (2011). An experience-based holistic account of the other-race face effect. In A.J. Calder, G. Rhodes, M.H. Johnson and J.V. Haxby (Eds.), *The Oxford handbook of face perception*. Oxford: Oxford University Press, 215-243.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61*, 81-88.

Schweinberger, S. R., & Burton, A. M. (2011). (Eds.). Person Perception 25 years after Bruce & Young (1986). *Special Issue of the British Journal of Psychology, 102*, 695-974.

Shepherd, J. W., Davies, G. M., & Ellis, H.D. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 105-132). London: Academic Press.

Shepherd, J. W., Deregowski, J. B., & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology, 9*, 205-212. doi: 10.1080/00207597408247104

Shepherd J. W., & Ellis H. D. (1973). The effect of attractiveness on recognition memory for faces. *American Journal of Psychology, 86,* 627-633. doi: 10.2307/1421948

Shepherd, J. W., & Ellis H. D. (1996). Face recall-Methods and problems. In S.L. Sporer, R.S. Malpass & G. Koehnken (Eds.), *Psychological issues in eyewitness identification*. Mahwah, NJ: Lawrence Erlbaum.

Shepherd, J. W., Ellis, H. D., & Davies, G. M. (1982). *Identification evidence: A psychological evaluation*. Aberdeen: Aberdeen University Press.

Solomon, C., Gibson, S., & Mist, J.M. (2013). Interactive evolutionary generation of facial composites for locating suspects in criminal investigations. *Applied Soft Computing, 7,* 3298–3306. doi: 10.1016/j.asoc.2013.02.010

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition, 127*, 105-118. doi: 10.1016/j.cognition.2012.12.001

Tanaka, J. W., & Farah, M. H. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A*, 225-245. doi: 10.1080/14640749308401045

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161-204. doi: 10.1080/14640749108400966

Valentine, T. (2001). Face-space models for face recognition. In M. J. Wenger & J. T. Townsend (Eds.) *Computational, geometric and process perspectives on facial recognition* (pp. 83-113). Mahwah, NJ: Erlbaum.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, 69, 1996-2019. doi: 10.1080/17470218.2014.990392

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modellng first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences, USA, 111*, E3353-E3361. doi: 10.1073/pnas.1409860111

Walsh, D., Oxburgh, G. E., Redlich, A., & Myklebust, T. (2016). *International developments and practices in investigative interviewing and interrogation*. London: Routledge.

White, D., Kemp, R.I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS ONE, 9(8)*, e103510. doi: 10.1371/journal.pone.0103510

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology: revised edition*. New York: Henry Holt.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145. doi: 10.1037/h0027474

Yin, R. K. (1970). Face recognition by brain-injured patients: a dissociable ability? *Neuropsychologia, 8*, 395-402. doi: 10.1016/0028-3932(70)90036-9

Young, A. W. (2011). Disorders of face perception. In A.J. Calder, G. Rhodes, M.H. Johnson and J.V. Haxby (Eds.), *The Oxford handbook of face perception*. Oxford: Oxford University Press, 77-91.

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959-974. doi: 10.1111/j.2044-8295.2011.02045.x

Young, A. W., & Burton, A. M. (in press). Recognizing faces. *Current Directions in Psychological Science*,

Young, A. W., & Ellis, H. D. (1976). An experimental investigation of developmental differences in ability to recognise faces presented to the left and right cerebral hemispheres. *Neuropsychologia, 14*, 495-498. doi: 10.1016/0028-3932(76)90078-6

Young, A. W., Hay, D .C., & Ellis, A. W. (1985a). The faces that launched a thousand slips: everyday difficulties and errors in recognizing people. *British Journal of Psychology, 76*, 495-523. doi: 10.1111/j.2044-8295.1985.tb01972.x

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985b). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*, 737-746. doi: 10.1068/p140737

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*, 747-759. doi: 10.1068/p160747

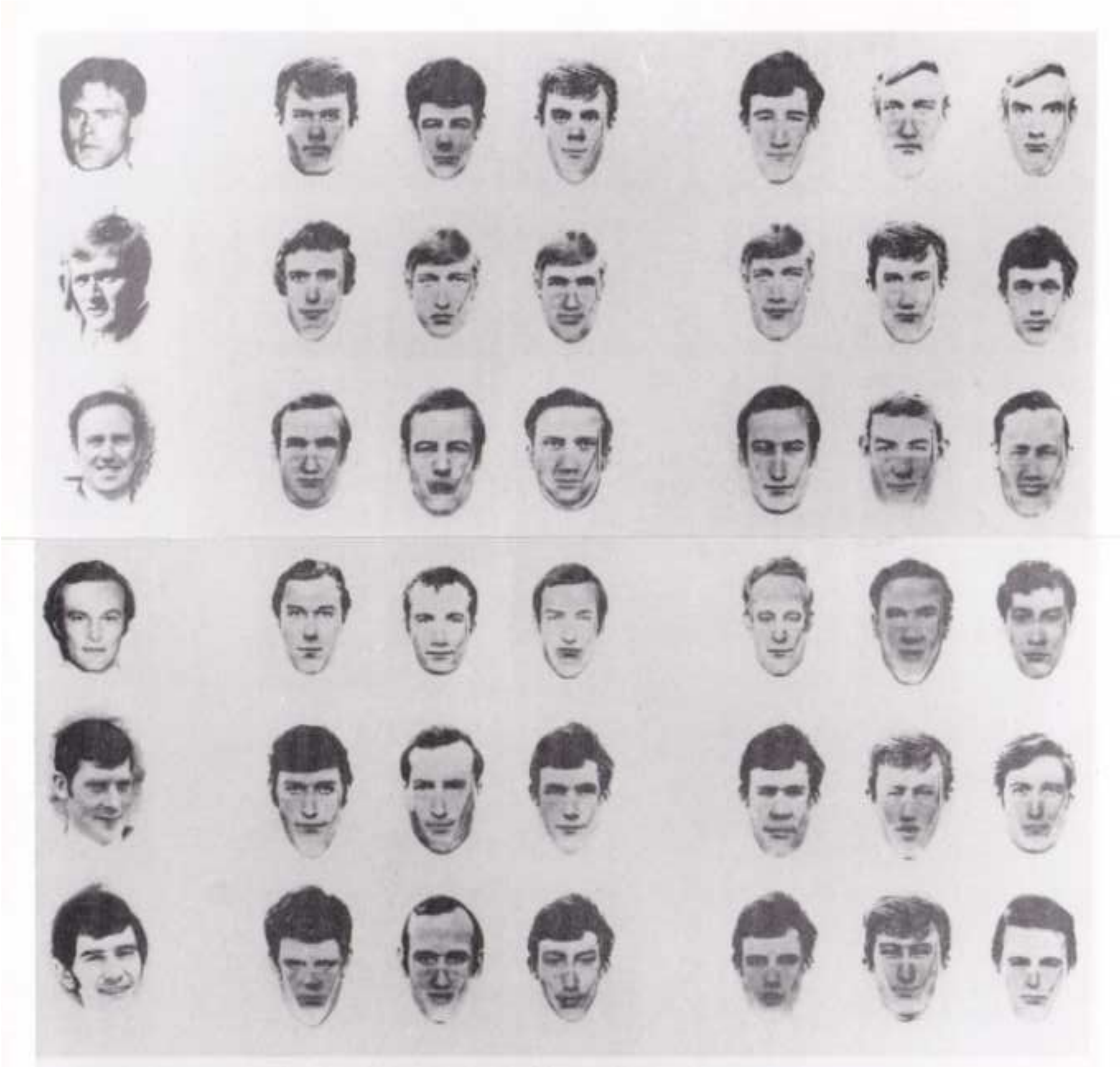
Figure Captions

Figure 1: Photofits of 6 different faces from Ellis, Shepherd and Davies (1975). Each face was seen briefly before an attempt to reconstruct it from memory. Each target face is shown in the left hand column, together with different participants' attempts to reconstruct the face using *Photofit* in the adjacent row. The first three reconstructions (columns 2-4 from the left) were made by participants considered relatively 'good' encoders and the other three reconstructions (columns 5-7) were made by participants considered relatively ‘poor’ encoders.

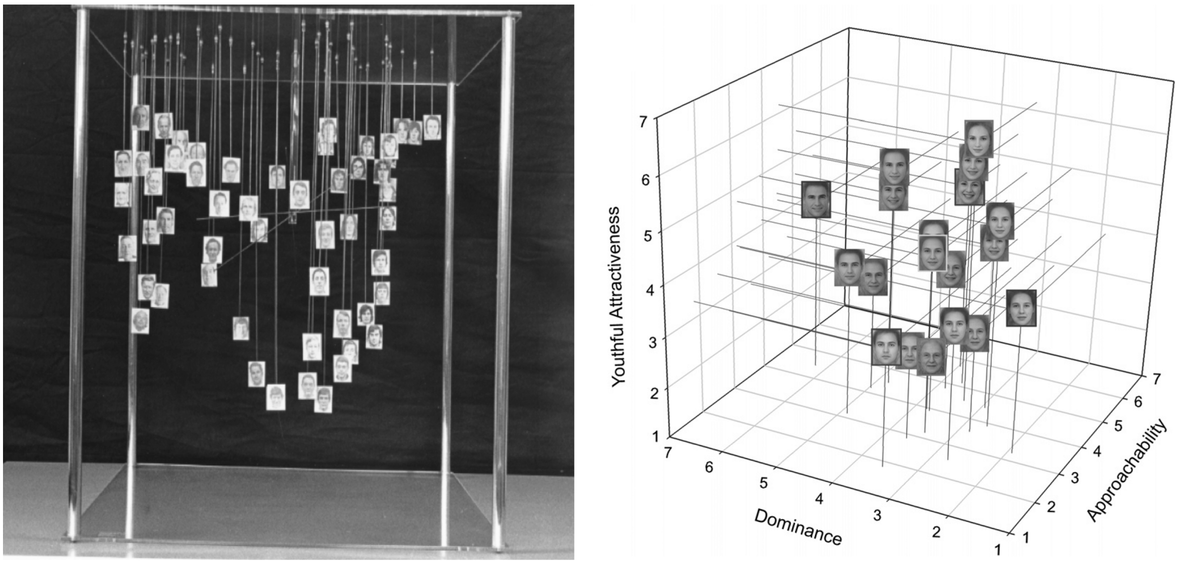
Figure 2: Left panel: Model of the three-dimensional face space for initial impressions of similarity in the physical appearance of male faces from the Aberdeen study, based on data analysed by the MINISSA multi-dimensional scaling program. Right panel: A related model of the three-dimensional face space for similarity in initial impressions of faces from a recent study by Sutherland et al. (2013), based on factor analysis of rated traits.

Figure 3: Percentages of participants correctly identifying and misidentifying target persons at 7, 30, 90 and 350 day intervals.

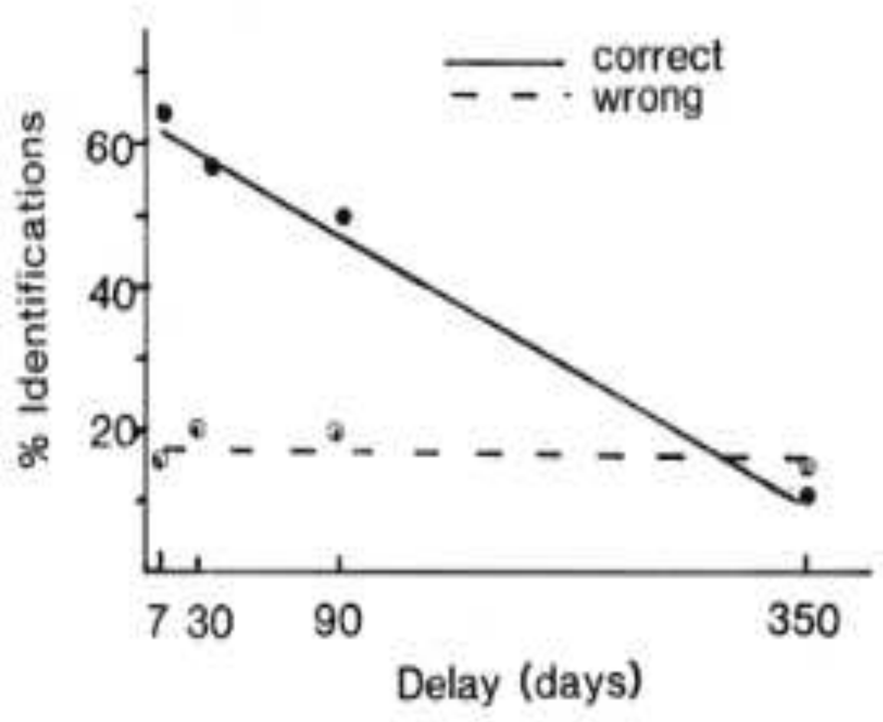
(Figure 1)



(Figure 2)



(Figure 3)


correct
wrong
% Identifications
60
40
20
7 30
90
350
Delay (days)