No face-like processing for objects-of-expertise in three behavioural tasks

Rachel Robbins and Elinor McKone

School of Psychology, Australian National University

(in press, Cognition)

Contact author:

Rachel Robbins
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, Ontario
L8S 4K1
CANADA
Ph: +1 905 525 9140 x 24761
Fax: +1 905 529 6225
Email: Dr.R.Robbins@gmail.com

Date: 24th Feburary, 2006

Short title: Face-specific cognitive processing

Abstract

In the debate between expertise and domain-specific explanations of "special" processing for faces, a common belief is that behavioural studies support the expertise hypothesis. The present article refutes this view, via a combination of new data and review. We tested dog experts with confirmed good individuation of exemplars of their breed-of-expertise. In all experiments, standard results were confirmed for faces. However, dog experts showed no face-like processing for dogs on three behavioural tasks (inversion; the *composite paradigm*; and sensitivity to contrast reversal). The lack of holistic/configural processing, indicated in the first two of these tests, is shown by review to be consistent rather than inconsistent with previous studies of objects-of-expertise.

Recognising faces is something people are very good at. It is well established that the process used to recognise another person's face differs from the process used to recognise nonface objects, at least when subjects have no specific expertise in the relevant object domain (e.g., Farah, 1996; Grill-Spector, Knouf, & Kanwisher, 2004; Moscovitch, Winocur, & Behrmann, 1997; Tanaka & Farah, 1993; Yin, 1969). Two main theories have emerged to explain this difference.

The domain-specificity hypothesis (e.g., Kanwisher, 2000; McKone & Kanwisher, 2005; Rhodes, Byatt, Michie, & Puce, 2004; Yin, 1969) suggests that the 'special' processing used for faces occurs only for faces. This hypothesis does not, per se, propose a mechanism for the origin of the special processing. It is possible, however, that special processing for faces has an innate component (de Haan, Humphreys, & Johnson, 2002; Morton & Johnson, 1991) and/or that it is necessary to obtain appropriate face experience at a particular time in development (e.g., a sensitive/critical period during infancy for the development of normal face processing, Le Grand, Mondloch, Maurer, & Brent, 2001, 2003, 2004).

In contrast, the expertise hypothesis (e.g., Carey, 1992; Diamond & Carey, 1986; Gauthier & Tarr, 1997; Meadows, 1974) suggests that 'special' processing for faces is a potentially generic ability that arises for faces because of substantial experience in individual-level discrimination; this predicts that the special processing can also arise for any other object class through the same mechanism (e.g., in expert dog show judges looking at dogs from their breed-of-expertise). An important assumption of the expertise hypothesis is that the period of life when this experience is obtained is irrelevant: object expertise can be developed entirely as an adult, and the predictor of processing style is merely the amount of practice. This assumption was made explicit by the original proposers of the hypothesis (e.g., Diamond & Carey, 1986; Carey, 1992), and has remained implicit in subsequent research, which tests for face-like processing in subjects who have in many cases obtained their expertise as adults or teenagers (e.g., Gauthier, Skudlarski, Gore & Anderson, 200; Gauthier & Tarr, 1997; Grill-Spector et al., 2004; Xu, 2005).

In terms of empirical tests of the different hypotheses, key neuropsychological results support domain specificity. In novices (i.e., subjects with no particular expertise in the object domain tested), there have been cases where face recognition is spared but object recognition is damaged (e.g., Moscovitch et al., 1997), and vice versa (e.g., De Renzi, 1986; McNeil & Warrington, 1993), showing a double dissociation between face and object recognition. In experts, this has been extended to a double dissociation between face processing and processing of objects-of-expertise. That is, individual patients have shown intact face processing with impaired recognition of objects-of-expertise (toy soldiers; Moscovitch et al., 1997), and also impaired face processing with intact processing of objects-of-expertise (brass instruments, Dixon, Desmarais, Gojmerac, Schweizer, & Bub, 2002; greebles, Duchaine, Dingle, Butterworth, & Nakayama, 2004; sheep, McNeil & Warrington, 1993). It is also possible for an object agnosic and prosopagnosic patient to become an expert with greebles despite not having re-learned expertise with faces after several years of exposure post-injury (Berhmann, Marotta, Gauthier, Tarr & McKeeff, 2005). These results, particularly the double dissociation, should not be possible according to the expertise hypothesis, which would have predicted that the normal versus impaired status of expert object processing should track that for face recognition, not dissociate from it.

However, studies using functional Magnetic Resonance Imaging (fMRI) have produced less clear-cut results. In novices, the Fusiform Face Area (FFA) shows a BOLD response approximately 2-3 times as strong to individual faces as to individual members of other object classes (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). In testing expertise effects, two results in the literature show <u>no</u> increase in activation in the FFA for objects in experts compared to novices (cars, Grill-Spector et al., 2004; cars, Gauthier, Skudlarski, Gore, & Anderson, 2000). Other findings, however, have indicated higher FFA activation in experts than novices (birds, Gauthier et al., 2000; birds and cars, Xu, 2005; lepidoptera: Rhodes et al., 2004), even if the level remains lower than that normally obtained for faces.

How should these results be understood? One interpretation is that, despite the neuropsychological evidence, the fMRI results should be taken as supporting the expertise hypothesis. An alternative view is that the small BOLD increase in fMRI studies represents something other than the brain carrying out the same type of computations for faces and objects-of-expertise. There are at least two reasonable candidates for an alternative explanation (e.g., Xu, 2005). One is that the increased signal is due to greater attention to the objects by the experts than the novices: the FFA produces some response to objects even in novices (e.g., Haxby et al., 2001), and attention is known to increase BOLD response (Wojciulik, Kanwisher, & Driver, 1998). According to this explanation, the attempts in most studies to equate level of attention to the objects between experts and novices, such as using one-back-matching tasks to force processing at the individual-exemplar level, are not fully successful. A second important consideration is that fMRI has limited spatial resolution, and the measured FFA may be combining the outputs of smaller areas with differing functions (as demonstrated in Schwarzlose, Baker, & Kanwisher, 2005). According to this interpretation, objects-of-expertise could appear to be activating the FFA when, in fact, greater resolution would reveal closely located but separate areas for faces and for objects-of-expertise.

Using behavioural studies to examine the nature of computations.

The present study addresses whether face-like computational processing emerges for objects-of-expertise. Given the difficulties of interpreting the fMRI data, this is a critical question in evaluating the expertise hypothesis.

The nature of the computations the brain performs with faces and objects-of-expertise is most cogently addressed by behavioural experiments. Currently, the common position in the literature is a presumption that behavioural studies support the expertise hypothesis. In the present article, we refute this position. We do so partly based on new results from three experiments conducted in dog experts, all of which show no evidence of face-like processing of dogs. Equally important, we include a careful review of previous data. This review shows that previous findings on standard tasks do not, in fact, support the expertise hypothesis.

Theoretically, the primary issue we are concerned with is whether experts develop the configural/holistic style of processing known to be 'special' to upright faces, or whether they continue to use the part-based processing style used for inverted faces and objects in novices.

The exact definition of the terms configural and part-based are a matter of ongoing debate (e.g., see Maurer, Le Grand, & Mondloch, 2002; Peterson & Rhodes, 2003; Tanaka & Sengco, 1997). Here, we take configural (or holistic) processing to mean strong integration of information from across multiple regions of the face at once, which occurs at a perceptual level, and includes processing of detailed spatial relational information between face features (e.g., nose-mouth distance) or perhaps between lower-order 'features' (e.g., distance from corner of eye to edge of nostril). We take part-based processing to mean independent local processing of individual features, where features are defined by obvious contour boundaries in the face.

Empirically, a number of standard tasks have been associated with configural processing. The details of these tasks, and their exact theoretical relationship to configural processing, are discussed in later sections of this article. Briefly, the phenomena we consider are: the disproportionately large inversion effect on memory for faces (Yin, 1969; see Experiment 1 for both our results and the review of previous findings), the Tanaka and Farah (1993) part-whole effect (see introduction to Experiment 2 for review), and the Young, Hellawell and Hay (1987) composite effect (see Experiment 2 for review and new results).

We also consider the sensitivity of face and object processing to contrast reversal (i.e., making the face look like a photographic negative; Experiment 3). We do so because of evidence that, in nonexperts at least, shape-from-shading information is particularly important in identifying faces (Galper, 1970; Kemp, McManus, & Pigott, 1990; Subramaniam & Biederman, 1997). In addition, a previous study (Gauthier, Williams, Tarr, & Tanaka, 1998) has considered contrast reversal relevant to testing the expertise hypothesis.

Testing the expertise hypothesis: What should be the objects, and who should be the experts?

According to the expertise hypothesis, face-like processing is most likely to emerge for objects when (a) the processing requirements of the task are the same for faces and objects, (b) the subjects are sufficiently expert in the object domain, and (c) the object class is matched to faces in important ways. In our own experiments, we satisfied the first of these criteria by using tasks that require discrimination of individual exemplars for both faces (e.g., Mary vs Jane) and the selected object class (e.g., Dog 1 vs Dog 2).

In judging what comprises sufficient expertise, we note that previous studies have used both experiment-trained subjects with approximately 8-10 hrs of training on individual members of the object class (the greeble studies, e.g., Gauthier & Tarr, 1997), and real world experts with years of experience (e.g., Diamond & Carey, 1986). Clearly, testing real world experts provides the best chance of identifying expertise effects. It is also important to provide experimental evidence that the expert subjects are, indeed, good at recognising their objects-of-expertise.

In terms of the type of expert, we wished to select experts who naturally discriminate their objects-of-expertise as individuals (Diamond & Carey, 1986). Not all experts do this; for example, bird experts usually discriminate birds at the species level, and car experts at the level of the model and year. Recent authors have suggested that the predictions of the expertise hypothesis should also apply to these experts who employ only subordinate-level categorisation (e.g., Gauthier et al., 2000; Xu, 2005). However, we chose experts who identify their objects-of-expertise at the individual level because these provide the closest match to the type of expertise people have with faces (and incidentally can provide the strongest evidence against the expertise hypothesis if null results are obtained).

In choosing an object class that is 'matched' to faces [1], Diamond and Carey (1986) suggested that two conditions are particularly important. First, the objects must share the same first order configuration; for faces, this comprises two eyes above a nose above a mouth. Second, individual exemplars of the class must differ only in second-order ways from this common configuration; faces, for example, differ in the exact distance between the eyes, or exact nose shape.

To these minimum criteria, our own experiments added the following properties. We wanted the objects to be natural rather than artificial, with individual members of the class differing because of genetic variability. Members of artificial object classes (e.g., greebles) tend to vary on only a small number of dimensions, and/or have differences located in a small number

[1] No object is perfectly matched to faces on all possible variables; if it was, it would be a face. Also note that we avoided objects with a face-like structure (e.g., animal faces, greeble heads). The question of whether expertise can 'stretch' the face recognition system's definition of a human face to cover animal faces (e.g., see Pascalis, de Haan, & Nelson, 2002) is a different one

of discrete locations. Natural objects are like faces, however, in that individual exemplars differ on many dimensions, with differences spread over the extent of the whole object. Finally, to examine contrast reversal effects, we wanted objects for which, like faces, part boundaries are often gradual, and shape-from-shading information is potentially useful.

These criteria are met by relatively few classes of objects, most obviously by animals (e.g., horses, cats, dogs). For practical reasons, we chose to test labrador dogs. Labradors are one of the most popular dog breeds in Australia, making it feasible to obtain experts. Labradors are also short-haired, avoiding the problem of clipping differences between individuals. Unlike dogs such as beagles, they have no strong colour boundaries, avoiding the problem of artificially encouraging a part-based processing strategy. Dog photographs are available in a standard 'show dog' pose (side on, usually with tail out), avoiding problems of substantial image differences between pictures of different individuals (a problem with cats); importantly dog-show judges and breeders are very familiar with pictures in this standard view (i.e., the side-on view we selected matches the particular expertise of the subjects). Finally, the breed standard for Labradors in Australia requires not only reference to local features of the dog (e.g., well-arched toes), but also to more global properties (e.g., strongly built, short coupled, very active, broad and deep through chest and ribs, broad and strong over loins and hindquarters), and our experts reported that they 'look at the whole dog' when judging quality, rather than merely check a list of local features.

Evidence of expertise in our subjects

Our experiments were designed to test whether labrador experts demonstrated face-like processing for their objects-of-expertise, on three behavioural tasks (inversion, composite task, contrast reversal). Two of these tasks also provided psychophysical data relevant to demonstrating expertise in our subjects. We now extract and summarise these data; full Methods for the tasks are described in Experiments 1 and 3.

Our subjects were 15 labrador show judges, breeders and trainers (aged 41-76 years). They had 5 to 42 years of experience with labradors, with a mean of 23.1 years. Twelve of the

---

from testing the generic expertise hypothesis that face-like processing can develop for objects of any structural form.

subjects had greater than 10 years of experience. This compares very favourably with previous studies testing real-world experts (e.g., Diamond & Carey, 1986; Gauthier et al., 2000; Rhodes et al., 2004; Tanaka & Curran, 2001; Tanaka & Taylor, 1991; Tanaka et al, 1996, cited in Tanaka & Gauthier, 1997). We also derived an estimate of total 'lifetime' number of labradors seen, using a time line questionnaire method to assist retrieval from different stages of life (as developed and validated in the drug-use field; Anglin, Hser, & Chou, 1993). Scores ranged between 43 individual labradors (in a guide dog trainer who had regular contact with each of these dogs over a year or more) and approximately 11635 individual labradors; the median was 2716 and the mean was 4553.

During testing, both anecdotal evidence and more formal evidence of expertise emerged. In an example of the former, subjects could sometimes name the breeder of a particular dog. Experts also commonly distinguished the dogs' country of origin. We had included Australian, British and American dogs, with experts finding the local dogs most attractive (indeed, a common comment was "Those American dogs are ugly!").

In terms of formal testing, only certain conditions from our experiments are relevant to assessing expertise. These conditions are those in which the dog stimuli were presented in their usual form – that is, upright, complete dogs, in normal contrast – as this is the appearance with which experts would have had the opportunity to develop expertise. Two experiments contained relevant conditions. In our recognition memory task, subjects learned one set of dogs, then later discriminated between the studied and the unstudied dog in a series of pairs (see Experiment 1). Table 1 shows percent correct for upright faces and dogs. Experts with more than 10 years' experience (N=12) were 11% more accurate for dogs than were control novices matched for age, sex and education ($\underline{t}$ (11) = 1.97, $\underline{p}$ < .04, for 1-tailed test of experts better than novices) and this did not reflect a speed-accuracy tradeoff as experts tended to be faster than age matched novices (see Table 4). In contrast, there was no difference between experts and matched novices for faces (1% difference, $\underline{t}$ < 1). More powerful data (more trials per subject) came from a same-different identity task with simultaneous presentation, where two images were briefly presented side-by-side showing either the same dog or two different dogs (see Experiment 3). Table 2 shows percent correct for upright faces and dogs (in normal contrast). In this case, data for experts and novices cannot be compared directly, because the older experts (age 41-72) were given a longer stimulus presentation time than the young adult novices (age 18-30 years), and because age-sex-education matched novices were not tested in this experiment. However, performance for faces and dogs can be compared within each subject group, because the same presentation time was used for both stimulus classes. The results show that experts (N=15) were as good with dogs as they were with faces (0% difference, $t < 1$), while young adult novices were 11% worse with dogs than faces, $t(19) = 4.87$, $p < .001$; there was also a highly significant interaction between stimulus class and expertise, $F(1, 33) = 24.08$, $p < .001$.Our finding that experts were as good with dogs as with faces in simultaneous matching is particularly strong: this argues that 5-42 years of experience has produced a level of perceptual expertise with dogs that is equivalent to that obtained with rather more years of experience with faces. (The same result was not obtained in recognition memory, but this task combines perceptual ability with memory factors.)

Overall, there is considerable evidence that the experts tested in our experiments were sufficiently expert to provide a valid test of the expertise hypothesis. According to the hypothesis, these particular subjects would clearly be predicted to show evidence of face-like processing for labradors.

Experiment 1 - Inversion effects on recognition memory

In object novices, all stimuli are remembered better upright than inverted, but this inversion effect is much larger for faces (usually 15-25%) than for objects (usually 0-8%; see Table 3). As pointed out by Valentine (1988), the disproportionate inversion effect for faces does not *per se* indicate configural processing for upright faces. In principle, the better performance with upright faces could arise from better part-based processing. For faces, the evidence showing that the large inversion effect arises from configural processing in the upright orientation, but only part-based processing in the inverted orientation, has come from other paradigms. These methods provide direct evidence of strong integration of parts into wholes for upright faces only (see introduction to Experiment 2). The relevant point for expertise studies is that, even if a study were to show an inversion effect for objects-of-expertise that was as large as for faces, it would be necessary to test other paradigms to know whether this reflected an emergence of configural processing, rather than some other form of advantage for upright.

In experts, previous inversion results are summarised in Table 3. The first line of the table gives the results from a classic study, in which Diamond and Carey (1986, Experiment 3) found that, for dog-show judges looking at dogs from their breed of expertise, the inversion effect on memory was much larger than in novices and, indeed, was as large as the experts' inversion effect for faces. This result was taken as compelling evidence for face-like processing of objects-of-expertise, and has been widely cited as such.

Surprisingly, however, in the 20 years since the publication of Diamond and Carey (1986), no replication of their finding has appeared in the literature. Indeed, the results of the relatively few studies that have re-tested inversion effects in experts do <u>not</u> show the same pattern as Diamond and Carey. As shown in Table 3, the results of the other studies are best summarised as showing a small increase, or no increase, in the inversion effect with expertise, with the effect for objects-of-expertise not approaching the size of that for faces.

In Experiment 1, therefore, we attempted to replicate the results of Diamond and Carey (1986, Experiment 3) using dog experts. Experts and novices were tested for long-term memory for both faces and dogs, shown upright and inverted. Our experiment was very similar to Diamond and Carey's in most ways. In particular, experts were shown dogs from their breed-of-expertise, and the level of real world experience was similar, particularly for our 12 most-expert experts, whose mean years' experience was 27.1 years (cf. 31 in Diamond & Carey’s Experiment 3).

However, our experiment differed from Diamond and Carey's in one critical way. A potential problem with their study is that, while the faces used as stimuli were novel to the subjects, it is not clear that this was true of the dog stimuli. Their dog pictures were taken from the archives of the American Kennel Club, and the experts were American Kennel Club judges. Thus, it is possible that experts had previously seen some of the champion dogs in person, and/or had previously seen some of the stimulus photographs stored in the archives. Any such exposure would presumably have been in the upright orientation. Under these conditions, pre-experimental familiarity with the dog stimuli could assist memory in the upright orientation, producing an artificially large inversion effect.

In the present experiment, we avoided this problem. Our subjects completed a post-test questionnaire that ensured the experts could not name, or give other individually identifying information about, any of the dogs used in our stimulus set.

Method

Subjects.

Labrador experts (N=15; age 41-76 yrs) were the people described earlier under "Evidence of Expertise in Our Subjects". The experts were located from lists of qualified judges obtained from the Australian National Kennel Council website, or from lists of local breeders obtained from the ACT Canine Association website [2]. Additional recommendations were also obtained from some of the dog experts (most of those recommended were also on the ANKC list). The only expert who was not a judge or breeder was a guide-dog trainer (guide-dogs in Australia are predominately labradors). Seven of the experts were male.

Given that memory performance tends to decline with age (e.g., Fastenau, Denburg, & Abeles, 1996), a second group of subjects was age-matched novices (N=15; age = 40 –76, mean = 58.7 years compared to the experts 58.2 years). These had no particular experience with dogs, and were matched to the experts on a one-to-one basis for age (± 2 years), sex, and level of education. They were contacted through advertisements placed around shopping centres, sporting clubs, in the newsletter of the University of the Third Age, and through friends and relatives of the experimenters. They were thanked with chocolates or a bottle of wine.

A third group was young adult novices (N=22, aged 18-43 with only two over 35, 11 males). These were undergraduate students of the Australian National University completing the

[2] http://www.ankc.aust.com/ and http://www.actca.asn.au/ respectively.

15 min experiment for course credit or $3. Most also participated in other unrelated experiments in a longer session.

All subjects were Caucasian (the same race as the faces). All reported normal or corrected-to-normal vision.

Design.

For each subject group, stimulus class (dogs vs. faces) and orientation (upright vs. inverted) were manipulated within subjects, and tested in a blocked fashion. For each stimulus class and orientation condition (e.g. upright dogs) there was a study (learning) phase, followed by a distracter phase, and then the memory test phase. For each condition, 15 exemplars were studied with 15 corresponding pairs presented at test. On each test trial, subjects saw two dogs (or two faces), one of which had appeared at study (old) and one of which was unstudied (new), and were asked to select the old one. The primary dependent measure was accuracy on this two alternative forced choice (2AFC) recognition memory task (chance = 50%). Reaction time was also measured.

Stimuli.

For the purposes of the 2AFC presentation, greyscale images of 60 yellow labradors (and 60 faces) were organised into 30 pairs, with appearance and stance matched as closely as possible across the exemplars in each pair, as shown in Figure 1. Note that, across all dogs, there was no one local feature that reliably distinguished old from new individuals: in any one test pair, one dog might have had its tail higher up than the other (e.g., middle pair in Figure 1), or one dog image might have more contrast then the other (e.g., top pair in Figure 1), but the subject also saw other images with a range of tail positions and contrast values.

Within each stimulus class, subjects received 15 pairs in the upright condition and the other 15 in the inverted condition, with the particular set in each orientation counterbalanced across subjects. Within each pair of dogs/faces, one of the two items was designated as studied while the other remained unstudied, and assignment to studied/unstudied status was counterbalanced across subjects. The same photograph of each dog/face was used at study and test, as (a) this was the procedure used in Diamond and Carey (1986), and (b) it was not possible to obtain two different photographs of each dog.

Dogs were taken from a mixture of sources including books (The Book of the Labrador Retriever, Nicholas, 1983; The Labrador Retriever Club of Victoria Inc.'s "Gold Book") and breeder web-sites in the public domain. There were 16 females and 44 males; half faced to the left and half to the right. Most of the dogs were American (approximately 38) or Australian (14) with the rest English or Canadian. As illustrated in Figure 1, there was a range of lighting direction in the photographs, as well as variability in image quality. Dog pictures were scaled to 4.9 cm – 6 cm from nose to tail (average 5.7 cm) by 3.5 cm - 4.6 cm from top of head to paws (average 4.2 cm). At the experimental viewing distance of approximately 45 cm, average images subtended visual angles of 7.2° horizontal and 5.3° vertical. Extraneous information was edited out of the image (e.g., grass stalks over paws; handler's fingertips on tail).

Face stimuli (Figure 2) showed Caucasians taken from a mixture of different online databases in the public domain (Stirling PICS: Nottingham-scans; University of Ljubljana CVL and CV, PTER, Velenje; Max-Plank Institute for Biological Cybernetics, Tuebingen, Germany), to match the variability in lighting direction and image quality in the dog stimulus set; there were also 10 females and 50 males. Photographs were front view with neutral expression. There were no glasses or beards, and small blemishes were edited out. Faces were cropped to exclude hair, but to retain as much forehead as possible, plus cheek and chin shape. This avoided memory being based on hairstyle. It also meant that, as with the dogs, the outline shape of the stimulus varied from face to face. After cropping, each face image was sized to 3.1 cm - 3.8 cm at the widest point (average 3.4 cm) by 4 cm - 4.6 cm at the tallest point (average 4.4 cm), corresponding to an average of 4.3° by 5.6° at 45 cm viewing distance.

All stimuli were placed on a neutral grey background. Brightness and contrast differences within each stimulus class set were minimised as far as possible, but there was still quite a lot of variation. Inverted stimuli were created by rotating the pictures 180°. All manipulations were done using Adobe Photoshop (5.5) software.

Procedure.

Pictures were presented on an iMac computer, with a 17” monitor, using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). Responses were recorded via the keyboard for young novices and via a NewMicros Button Box connected to the iMac via a Keyspan USB twin serial adaptor for the experts and matched novices.

During each study phase, dogs or faces were presented one at a time in the centre of the screen. Subjects were instructed to learn these for a later memory test. Each picture was presented for 5000 ms, with an intertrial interval of 575 ms. Presentation order was randomised for each subject.

During each distracter phase, subjects were presented with multiplication problems on the screen and were instructed to answer, on paper, as many as they could in 1 min. No subject completed all the problems in this time.

In each memory test phase, the two stimuli presented on each trial were shown simultaneously, at the same height, and 13.3 cm (16.8°) apart centre to centre. The stimuli remained on the screen until the subject responded. Subjects pressed one key to indicate that the left exemplar was “old” and another to indicate that the right exemplar was “old”. The old exemplar appeared on the left and right equally often. The intertrial interval was approximately 100 ms. Presentation order of the pairs was randomised for each subject.

Across subjects, the order of testing the four conditions (faces upright, faces inverted, dogs upright, dogs inverted) was counterbalanced. The order of conditions for a given age-matched control was the same as for their particular paired expert.

Structure of the testing sessions for experts, and post-test questionnaire.

Experts participated in all three of our experiments. A first 1-1.5 hr session included Experiments 1 and 3, and a second 1-1.5 hr session included Experiment 2. Experts were thanked with a book voucher or a bottle of quality wine.

The post-test questionnaire was given after all three experiments were completed. For this, the expert subjects were shown each dog stimulus used in the experiment, one at a time, in the upright orientation. Subjects were asked to (a) name the dog and/or give other identifying information if possible (e.g., 'awarded the Wilnora Show Trophy as Top Show Dog, 2005'), and (b) provide any other information they could about the dog (e.g., breeder, country of origin, good/bad example of breed).

Results

Confirming unfamiliarity with the dog stimuli.

On the post-test questionnaire, of the 15 dog experts, only 5 gave specific information about any dog, all of which was incorrect. None correctly named or identified any dog, although four each misnamed one dog as another which did appear in the experiments.

Inversion effects for each subject group.

Data from the young novices were expected to confirm standard findings (e.g., Yin, 1969), namely that, when subjects are not experts in the object domain, there is a much larger inversion effect for faces than for objects. Figure 3a shows this expected pattern. The difference between upright and inverted was significant for both faces, $t(21) = 5.10$, $p < .001$, and dogs, $t(21) = 2.41$, $p < .05$. An interaction between stimulus class and orientation, $F(1, 21) = 5.74$, $MSE = 186.01$, $p < .05$, confirmed the inversion effect was significantly bigger for faces (21%) than for dogs (7%).

Figure 3b shows results for the age-matched novice group. Overall, these novices performed slightly worse than our young novices, consistent with a small decline in memory with age. More importantly, this novice group again showed a larger inversion effect for faces (24%) than dogs (3%). The interaction between class and orientation was again significant, $F(1, 14) = 11.37$, $MSE = 149.11$, $p < .01$, although in this case, the inversion decrement was significant only for faces, $t(14) = 5.22$, $p < .001$, and not for dogs, $t < 1$ (possibly in part due to the smaller N than for young novices).

For experts, Figure 3c reveals an interesting result. The inversion effect for faces was large (21%), $t(14) = 4.69$, $p < .001$, the inversion effect for dogs was small (7%) and nonsignificant, $t(14) = 1.62$, $p > .05$, and there was an interaction between class and orientation,

$F(1, 14) = 6.95$, $MSE = 102.34$, $p < .05$. This fails to replicate Diamond and Carey's (1986, Experiment 3) finding of a face-sized inversion effect for dogs in dog experts. Indeed, results for experts showed a pattern indistinguishable from that for novices.

Statistics comparing patterns for experts and novices.

Separate ANOVAs compared results for experts with those for age-matched novices, and young novices, in turn. These tests each showed no interaction between stimulus class, orientation and expertise ($F(1, 14) = 1.58$, $MSE = 66.59$, $p > .2$ for matched novices, and $F < 1$ for young novices). That is, there was no evidence that experts showed a dog-specific increase in the size of the inversion effect, as would be predicted by the expertise hypothesis.

Reaction time data.

Analyses so far have considered only accuracy. Table 4 shows reaction time (RT) data for the three groups. As can be seen, RT results did not in any way contradict the accuracy results or indicate speed-accuracy trade-offs. Statistical analyses of RT data was less powerful than that of the accuracy data (based on only 15 trials per subject, RTs are highly variable) but, apart from non-significance of some trends that had been significant in the accuracy data, agreed with the accuracy findings. That is, all groups were faster for upright stimuli than inverted stimuli, and showed at least a trend towards larger inversion effects for faces than dogs, with no effect of expertise on the size of the inversion effect for dogs.

Correlations with amount of experience for experts.

A final analysis approached the test of expertise effects in a different way. Support for the expertise hypothesis would be obtained if, within the experts group, the size of the inversion effect increased with greater experience with the upright orientation. However, this result was not obtained. There was no correlation between the size of the inversion effect (% correct upright – % correct inverted) and either years of experience ($r = .096$, $p > .5$; Figure 4) or estimated number of dogs seen in lifetime ($r = .324$, $p > .2$).

Discussion

Our results have shown that, in contradiction to Diamond and Carey (1986), the inversion effect for dogs in experts (7%) remains much smaller than that for faces (21%). Indeed, our experiment has produced one of the weakest expertise results in the literature. The dog inversion effect for experts (7%) was at best slightly bigger than for age-matched controls (3%), and correlated slightly with lifetime number of dogs seen ($\underline{r} = .32$), but was no bigger than the dog inversion effect for young novices (7%) and did not correlate at all with years of experience ($\underline{r} < .1$).

Across all studies except Diamond and Carey (1986), the clear findings (Table 3; present results) are that expertise produces at most a small increase in the size of the inversion effect for objects, which does not approach the size of the effect for faces, even with extensive real world experience. It is also worth noting that the exact size of the small expertise effect in any one experiment is quite variable, ranging from no change in the inversion effect between novices and experts (Xu, Liu, & Kanwisher, 2005, accuracy measure), through a nonsignificant but possible small increase (present study; Bruyer & Crispeels, 1992; and probably Gauthier et al., 2000, although the relevant statistical test was not reported), to a significant increase in one study (Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002). A likely reason for some of this variability across studies is that, in the experiments testing long-term memory at least, the number of trials one can test per subject is small (otherwise, memory approaches floor).

This leaves Diamond and Carey (1986) as a stand-alone result. Our explanation of their unusual finding is that theirs was the only study in which expert subjects might have been familiar with the dog stimuli prior to the experiment. Although it is difficult to know after 20 years whether this was the case, it could explain a large inversion effect for dogs because experts would be familiar with those dogs specifically in the upright orientation. This pre-experimental familiarity would assist their memory in comparison to inverted dogs, and in comparison to unfamiliar faces.

In conclusion, findings from our own study in conjunction with previous research suggests that inversion effects increase only slightly with expertise. The common view that inversion effects become face-like for objects-of-expertise is not supported.

Experiment 2 - Configural processing and the composite effect

Experiment 1 failed to show face-like processing for objects-of-expertise on a recognition memory task. Note again that, even had it done so, this would not demonstrate directly that configural processing had occurred for objects of expertise: improvements with experience for upright could come from other sources (e.g., improved part-based processing).

In demonstrating configural processing for faces, two paradigms have become standard (e.g., see review by Maurer et al., 2002). In Tanaka and Farah's (1993) part-whole paradigm, the subject learns a whole face (e.g., Mary) and is later asked to select the old stimulus from a pair in either a part-alone condition (Mary's mouth vs. Jane's mouth) or a part-in-whole condition (Mary's mouth in Mary's face vs. Jane's mouth in Mary's face). Memory is better in the whole condition than the part condition for upright faces, but not for inverted faces, corresponding to the perception that, in upright faces, changing one feature of the face changes the appearance of other regions (e.g., replacing a large mouth with a small one can make the eye-nose region seem more squished up). The part-whole paradigm, and others very like it (e.g., the 'complete-probe advantage'; Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999), have been fairly extensively tested for objects. In novices, the standard result is that the whole-over-part advantage is large for faces, but small (e.g., Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999; Gauthier & Tarr, 1997; Tanaka et al. 1996, cited in Tanaka & Gauthier, 1997) or absent (Tanaka & Farah, 1993) for objects.

In experts, the results of the part-whole paradigm are also clear-cut. A good range of object classes and expert types has been tested, and the size of the part-whole effect for objects does <u>not</u> increase with expertise. Relevant data are shown in Table 5. This fails to support the expertise hypothesis.

It is clear that there is no effect of expertise on configural processing as measured by the part-whole task. This conclusion has been accepted by the authors of the original studies (see Tanaka & Gauthier, 1997, p. 111; Gauthier & Tarr, 2002, p 432). However, Gauthier and Tarr (2002) then argued that the part-whole task does not assess 'real' configural processing; they suggested instead that it assesses only an advantage of context that is generic to all stimuli. It is also possible that the task is affected by transfer-appropriate processing in memory, in that the usual whole-over-part advantage is reversed when parts, rather than wholes, are learned at study (Leder & Carbon, 2005). However, we disagree with the idea that the task does not tap configural processing at all, because this fails to explain why the part-whole effect is larger for faces than for objects: instead, we suspect that the part-whole effect combines two components, a small generic context effect (as also occurs in the word superiority effect; Reicher, 1969; Wheeler, 1970) and/or transfer-appropriate processing effect, and a larger effect limited to faces that arises from true perceptual integration.

Rather than engage in lengthy theoretical debate, this conflict can be empirically addressed by testing other tasks that measure configural processing. In Experiment 2, therefore, we employed Young et al.'s (1987) <u>composite task</u>, which has been widely accepted as a strong method for assessing configural processing for faces.

The original version of the composite paradigm tested familiar faces (Young et al., 1987). The top half of one face (e.g., George Bush) was combined with the bottom half of another (e.g., Tony Blair) to make a composite. The halves were then presented in either an <u>aligned</u> condition, or in an <u>unaligned</u> condition where one half was horizontally offset by approximately half a face width. The task was to name a specified half of the face (e.g., the top half). For upright faces, subjects were substantially slower to name the half-face in the aligned stimuli than in the unaligned stimuli (Young et al, 1987; also see Carey & Diamond, 1994; Robbins & McKone, 2003). In interpreting this result, it is important to note that the two conditions compared are matched for simple response competition from the nontarget half (e.g., if naming George Bush, the bottom half suggests 'Tony Blair' in both the aligned and unaligned conditions). Thus, the interference effect observed in the aligned condition can only come from the perceptual integration that makes the two old halves appear to merge to make a 'new person', providing very direct evidence of configural processing. As would also be expected, inverted faces show no composite effect (Young et al., 1987; Carey & Diamond, 1994; Robbins & McKone, 2003), indicating independent processing of the two halves even when they are aligned, consistent with part-based processing.

Turning to objects, only two previous studies (Gauthier & Tarr, 2002; Gauthier et al., 1998) have assessed the composite effect for objects-of-expertise. Indeed, only one has tested the effect even in novices (Gauthier & Tarr, 2002). Results are shown in Table 6. Here, we have presented both accuracy and reaction time differences as positive if they were in the correct direction for a composite effect (i.e., aligned less accurate than unaligned; aligned slower than unaligned), and negative if they were in the reverse direction. For greeble novices, Gauthier and Tarr (2002) found no composite effect. For greeble experts, the pattern across a number of tests is also of no composite effect. Surprisingly, these results have been cited (e.g., Tarr, 2003) as showing face-like processing for objects of expertise. However, it is apparent from the table that this interpretation is only possible with selective reading of the data: that is, by focussing on the one condition that shows a close-to-significant composite effect in experts (+115 ms) while ignoring the null and reverse effects that were found in other conditions.

Although the composite effect has not been found for greebles in experiment-trained experts, it remains possible that the effect might emerge for natural objects in real world experts. Dog experts have many more years of experience than greeble experts, in a wider range of settings, and dogs have more sources of variation across individuals than greebles (meaning that configural processing might be more important). Thus, in Experiment 2 we examined the composite effect in our labrador experts.

To do so, it was necessary to modify the original Young et al. paradigm somewhat, to allow testing of novel rather than familiar stimuli. Our method was similar to that recently reported by Le Grand et al. (2004; also see Hole, 1994) for faces. We simultaneously presented a pair of composite faces, or a pair of composite dogs, on each trial. Each composite was formed from two different individuals and, as shown in Figure 5 (faces) and Figure 6 (dogs) both composites in the pair had their halves either aligned or unaligned. The task was to decide whether a specified half (top or bottom) was the same or different in identity across the pair members (the half-to-ignore was always of different identity). Each pair of stimuli was presented for a limited time with the intention of measuring accuracy rather than reaction time [3]. The presentation time used (600 ms for young novices, 600-850 ms for experts, depending on age) was chosen to allow subjects to look at each exemplar 1-2 times; this minimised opportunities for feature-by-feature comparison.

Our question of interest was whether the comparison of aligned and unaligned conditions produced any evidence of configural processing for dog experts, looking at upright dogs. We also tested inverted dogs, and upright and inverted faces, expecting to confirm a composite effect for upright faces in both dog experts and dog novices, but no composite effect for inverted faces or inverted dogs in either subject group.

Before describing the experiment in detail, it is important to note that, in the same-different identity task, only the results of same trials are theoretically meaningful (Le Grand et al., 2004; Hole, 1994). This contrasts with the approach taken by one previous study (Gauthier & Tarr, 2002) where same and different trials were combined to give an overall d' measure. Unfortunately, this is not a valid procedure, because the direction of effect corresponding to configural processing can be reliably predicted only for same trials. For same-identity trials, the direction corresponding to configural processing matches that found in Young et al. (1987).

---

[3] Indeed, we did not consider reaction times a valid measure. In studies designed to use reaction time as the measure, the stimulus normally stays on the screen until the subject responds (as in Experiment 1). Under these circumstances, the subject gains more stimulus information by waiting longer to respond, and thus reaction time is a meaningful measure of processing difficulty. However, when brief presentation is used (as in Experiment 2 and Experiment 3) this is not necessarily the case. In our experience, on trials where subjects feel they are guessing after the stimulus has disappeared, different subjects take different strategies. Some agonise about the decision for quite some time; others respond rapidly on the grounds that they might as well guess now, because no further information to help their decision will be forthcoming. Thus, a difficult decision can be associated with either long reaction times or short reaction times, making reaction time an invalid measure of processing difficulty. (In response to any readers still concerned about possible effects on reaction times: analysis of RTs showed no composite effect for experts on upright dogs.)

Specifically, the aligned condition should always be less accurate than the unaligned condition, for the following reason. When the top half of the same individual is paired with two different bottom halves, any perceptual integration between aligned halves will make the tops appear less similar to each other than they really are, making it harder to say ‘same’ (see Figures 5 & 6). Now consider different-identity trials. Here, the prediction corresponding to configural processing is intrinsically unclear. When responding to the top half, configural processing could make the aligned condition more accurate than the unaligned condition for some stimuli (i.e. the reverse prediction to same trials), if the two bottom halves happen to be quite different in appearance (the additional dissimilarity should assist a correct ‘different’ judgement). On the other hand, configural processing could also make the aligned condition less accurate that the unaligned condition, if the two bottom halves happen to be fairly similar in appearance (the enhanced similarity will impair a ‘different’ decision). Thus different trials (and d') are only of use in assessing overall bias effects, not in evaluating the composite effect.

Method

Subjects.

Subject groups were the dog experts tested in Experiment 1 (N=15), and a new set of young novices (N=23, ages 18-30, approximately half male). Young novices were Caucasian undergraduates reporting normal or corrected-to-normal vision; 8 received course credit, 15 received $10. Thirteen had previously participated in the task presented as Experiment 3. No age-sex-education-matched novices were tested for this experiment because it took nine months to find them for Experiment 1 (also, the outcome of the experiment made data from such a group largely irrelevant; see Results).

Design.

A simultaneous presentation, same-different version of Young et al.'s (1987) composite paradigm was used. Each composite stimulus was created by taking the top half of one face (dog) and combining it with the bottom half of a different face (dog). A horizontal split was chosen for dogs because it disrupts a number of the key global properties in the relevant labrador standard (including, for example, 'broad and deep through chest and ribs'; 'broad and strong over loins and hindquarters'). For aligned composites, the two halves were aligned. For unaligned composites, the two halves were offset horizontally by approximately a quarter of the width of the face (dog). On each trial, two composites were presented simultaneously, either both aligned or both unaligned, and the subject indicated whether a given half of the two stimuli (e.g., the top half) was the same or different in identity. The half-to-be-ignored (e.g., the bottom half) was always different-identity for the two stimuli. In instructions to subjects, halves were referred to as ‘forehead’ versus ‘chin’ for faces, and ‘head’ versus ‘legs’ for dogs, to avoid confusion with ‘top’ and ‘bottom’ when the stimuli were presented inverted.

Class (faces vs. dogs), orientation (upright vs. inverted), half-to-compare (forehead vs. chin for faces, head vs. legs for dogs) and alignment (aligned vs. unaligned) were all manipulated within subjects. Class, orientation and half-to-compare were blocked; for example, one block of trials presented dogs inverted for matching of the legs half. On each trial, the composite pairs were presented for a brief period (600 ms for novices, 600-850 for experts) and accuracy of same-different responses was the dependent measure. Instructions were to judge identity with any changes in size or brightness/contrast to be ignored.

Stimuli.

Twelve dogs and twelve faces were chosen from the larger stimulus sets used for Experiment 1. Each class was divided into two sets of six exemplars. A given subject saw one set of exemplars for matching top halves and the other for matching bottom halves, with assignment of sets to half-to-compare condition counterbalanced across subjects. Stimuli within each set of six exemplars were chosen to satisfy several criteria: (1) they were all the same sex (all male); (2) they were chosen from only one book/database so that quality of photographs was similar across the set; and (3) when cut in half, any half-exemplar could be made to fit with any other half-exemplar with no distortion of the shape.

Forming composites. To make halves, each face was cut just below the eyes, and each dog was cut from mid-chest to just below the tail. Within a set of six exemplars, 30 composites were then formed by combining the top half of each exemplar with the bottom half of all other exemplars (6 top halves each combined with 5 possible bottom halves = 30 composites). Two versions of each of the 30 composites were created. In one version, the two halves were made to join up reasonably neatly by keeping the bottom halves unaltered and adjusting the size and brightness/contrast of the top half. These top-half-altered versions were used in blocks requiring matching of the top half, forcing subjects to base their same-different decision for a pair of composites on whether the top half had the same identity, rather than simply matching low level visual information. In the other version of the 30 composites, the two halves were made to join up by keeping the top halves unaltered and adjusting the size and contrast of the bottom half. These bottom-half-altered versions were used in blocks requiring matching of the bottom half. Finally, the 30 composites of each version in the aligned format were then also created in the unaligned format, in both a left-offset format (i.e. the bottom half shifted to the left) and a right-offset format (i.e., the bottom half shifted to the right).

Forming pairs of composites. To form pairs of composites for simultaneous presentation, the following procedure was used. For each set of 30 composites (derived from 6 exemplars), all 60 possible same pairs were created (for a top-half-to-compare block these include: top1-bottom2 composite paired with top1-bottom3 composite; top1-bottom2 composite paired with top1-bottom4 composite; etc). All possible different pairs were also created. With the constraints that the top half must be different across the pair, the bottom half must be different across the pair, and that the two halves of a given original face (dog) should not appear on the screen simultaneously (i.e., as the top half of one composite and the bottom half of the other), there exist 180 such combinations for each set of 30 composites.

Structure within each block. Subjects performed one block of 240 trials for each class x orientation x half-to-compare condition (e.g. dogs inverted legs-to-compare). Order of testing these conditions was counterbalanced across subjects. Each block included an equal number of aligned and unaligned trials, presented in a different random order for each subject. Within a block, each subject saw the 60 same trials for the assigned set of exemplars, plus 60 different trials selected randomly from the 180 possible such trials. These were all in the aligned format. For every aligned composite trial, there was another trial that presented exactly the same composite items in the unaligned format, for the total of 240 trials per block. Unaligned composites shown together always had offsets in the same direction (half of trials were left-offset, half right-offset), with the particular items assigned to each offset direction chosen randomly for each subject. Once a given subject's trial assignment had been determined for a given class and half-to-compare in the upright orientation (e.g., for dogs upright legs-to-compare), the trial structure of the inverted orientation was exactly matched to this. That is, all stimuli were exactly the same except rotated by 180° (and presented in a new random order).

Target-half cues. So that subjects did not forget which half to compare part-way through a block, all trials showed two short horizontal lines either above the stimuli (match the part on the upper half of the screen) or below the stimuli (match the part on the lower half of the screen). These indicator lines were set to the side so as to be visible in peripheral vision but not to interfere with processing of the composites.

Jittering of position. From trial to trial, the position of each composite on the screen was jittered (following Hole, 1994). Because half-to-compare was blocked, jittering was particularly important to avoid pre-focusing of an attentional 'spotlight' on the region of space where one half would appear; such an attentional window could work against the emergence of a composite effect, as it could allow tuning out of the half-to-ignore at a very early stage of visual processing (e.g., in area V1) such that higher-level visual recognition systems never received information from the nontarget half. Each composite stimulus of a pair appeared in one of 5 positions. Horizontal distance between the two composites (centre to centre) varied from 8 cm (10.2°) to 11 cm (13.9°) for faces and 7.5 cm (9.5°) to 8 cm (10.2°) for dogs. Vertical distance varied from 2 mm (0.3°) to 1 cm (1.3°) for faces and 2 mm (0.3°) to 5 mm (0.6°) for dogs. Position was chosen randomly on each trial, with the constraint that the two composites were never presented at the same height.

Procedure.

Apparatus was as for experts in Experiment 1. Initial instructions to subjects included a step-by-step talk through of a same trial and a different trial, using exemplars not from the main experiment. Subjects were then warned that the stimuli on each trial would be available for only a brief period, and given 14 practice trials. They responded same-identity via one button and different-identity via another.

Presentation duration was always the same for faces and dogs, and for upright and inverted. For young novices, the pair of composites on each trial was presented for 600 ms; stimuli were far enough apart that eye movements were required to foveate each item, and in pilot testing young subjects generally reported that 600 ms allowed 1-2 fixations per item. The pair of composites was followed by a blank screen until the subject responded. If the subject responded before 600 ms (this rarely happened), the stimulus was removed. The next trial began 200-400 ms after response. For experts, presentation time was increased to keep accuracy off floor, and to approximately match overall performance levels to that of the novices so that the size of any composite effect could be compared across the two groups. In early testing, some experts reported that they had trouble making fast enough eye movements to see both stimuli in a pair in 600 ms. For each expert, presentation time was adjusted in the practice phase until they felt that they could look at each stimulus at least once. Times per trial ranged from 600-850 ms with longer times generally used for older experts.

Pilot testing had revealed a general bias to respond same, presumably because this is the default response if no difference between the two stimuli was found within the brief presentation time. Thus, subjects were explicitly informed that the correct answer would be same on 50% of trials and different on 50% of trials. This reduced bias, although it did not fully eliminate it, for the young novices. Experts were more problematic. After testing 12 experts, it was noticed that some had a strong bias to respond same rather than different, meaning that scores in some of the same-identity conditions of interest were approaching ceiling for some experts. Two methods to overcome this were tried. The first was to give experts even longer to view the stimuli. This was used with one subject (850 ms) but did not remove bias. A version of the experiment was then created which included feedback (a beep on incorrect responses). This was pilot tested with five novices similar to those described above; results showed less bias than for the original novices. Thus, the last two experts were tested with auditory feedback. To avoid increases in overall accuracy, time for the last two experts was reduced again to 650 ms and 600 ms respectively. Note that the minor procedural differences between different subjects do not affect the comparisons of interest (i.e., aligned vs. unaligned, faces vs. dogs, upright vs. inverted) as all these were within-subjects, where exactly the same procedure was used for each class of stimulus.

Results

For faces, all results were very similar for top-half-to-match and bottom-half-to match blocks. For dogs, overall accuracy was slightly higher for top-half-to-match, but the pattern of composite effects across conditions was similar for the two halves. Scores were therefore collapsed over half to maximise power. Figure 7 shows accuracy on same trials, as used to assess the composite effect. Data for different trials are provided in Table 7; note again that these are relevant only to assessing general bias effects.

Young adult novices.

For the novices, results were as expected. As shown in Figure 7a/b, there was a highly significant composite effect for upright faces, $t(22) = 4.85$, $p < .001$ [4], but none for inverted faces, $t < 1$, upright dogs, $t < 1$, or inverted dogs, $t < 1$. These differences in pattern were confirmed via a 3-way interaction between stimulus class, orientation, and aligned versus unaligned trials, $F(1, 22) = 6.15$, $MSE = 6.64$, $p < .05$. Faces showed the expected 2-way interaction between alignment and orientation, $F(1,22)=30.87$, $MSE=2.70$, $p<.001$, while dogs showed no alignment by orientation interaction, $F<1$, $MSE=8.78$, and also no main effects of alignment, $F<1$, $MSE=15.28$ or orientation, $F(1,22)=1.36$, $MSE=38.11$, $p>.25$.

---

[4] The absolute size of this effect (i.e., aligned-unaligned difference) was smaller in our study than in Le Grand et al. (2004). The basic method of the two studies was similar, but likely factors in the smaller effect here are (a) Le Grand et al.'s stimuli had less visible joins between the two halves than ours, and (b) they were able to use each face only once in the entire stimulus set (Richard Le Grand, personal communication, October 2004). Our procedures for faces were constrained by the stimuli available for dogs.

Composite effect for experts.

For faces, experts (Figure 7c/d) showed the expected results of a strong composite effect for upright faces, $t(14) = 3.24$, $p < .01$, but none for inverted faces, $t < 1$, with a 2-way interaction between alignment and orientation, $F(1, 14) = 7.54$, $MSE = 23.60$, $p < .02$. Most interestingly, for dogs there was no composite effect for either upright dogs, $t < 1$, or inverted dogs, $t(14) = 1.44$, $p > .15$. There was also no alignment x orientation interaction, $F < 1$, and no main effects of either variable, Fs < 1. Although the 3-way interaction between stimulus class, orientation, and aligned versus unaligned trials was not significant, $F(1, 14) = 1.75$, $MSE = 28.04$, $p > .2$, there was a significant 2-way interaction between orientation and alignment, $F(1, 14) = 12.43$, $MSE = 11.33$, $p < .01$, and a 2-way interaction between class and alignment that approached significance, $F(1, 14) = 3.78$, $MSE = 15.93$, $p = .072$.

In Figure 7d, a possible concern with the dog results was that experts' mean performance on same trials was closer to ceiling than might be desired. To demonstrate that this was not hiding a composite effect, we also analysed data for the sub-set of experts whose accuracy was 90% or less on the unaligned upright dogs condition (noting that any composite effect would show up as even lower performance in the aligned condition). This subset (N=10) showed less of a ceiling effect for dogs than was present in the full data set for faces. However, results did not change: there was no composite effect for upright dogs (aligned = 84.5%, unaligned = 84.7%, $t < 1$), while the composite effect for upright faces was still present, $t(9) = 2.49$, $p < .05$. Thus, the finding of no composite effect for dog experts looking at dogs cannot be attributed to ceiling effects.

Statistics comparing patterns in experts and novices.

For dogs, a mixed ANOVA with alignment (aligned vs unaligned) and orientation (upright vs inverted) as within-class factors, and experience (novices vs experts) as a between-class factor was conducted. This showed that there were no interactions with experience, all ps > .11. For faces, formal comparisons between experts and novices were not conducted because of large differences in variance across the groups; however, as both groups showed all the expected effects this was not considered problematic.

Note that the novices in this experiment were all young; an age-matched control group was not tested. However, given that both young novices and older experts showed no composite effect for dogs, it seems unlikely that an older novice group would show anything other than the same result.

Correlations with amount of experience.

The size of the composite effect for upright dogs (unaligned – aligned) did not correlate with years of experience, $r = -.48$, $p = .069$ (Figure 8), or number of dogs seen, $r = -.48$, $p = .070$. Indeed, in both cases the trend was towards a weaker composite effect with greater expertise, which is in the opposite direction to that predicted by the expertise hypothesis.

Discussion.

Experiment 2 revealed no composite effect for upright dogs in dog experts, despite a highly significant effect for upright faces in the same subjects using exactly the same procedure (i.e., presentation times, auditory feedback). There was also no correlation between the size of the composite effect and either measure of amount of expertise, and no difference between experts and novices in the pattern of composite effects across stimulus types (i.e., both groups showed a composite effect for upright faces, and both groups showed none for inverted faces, upright dogs, or inverted dogs). Thus, there was no face-like processing of dogs in dog experts. This result is consistent with the lack of composite effect in greeble experts (Table 6), but extends the result to natural objects, and the stronger case of experts with many years of real-world experience.

Overall, results of the present study and of previous investigations again fail to support the expertise hypothesis. In particular, both the composite and part-whole tasks fail to support the critical prediction of configural processing for objects-of-expertise. [5]

Experiment 3 - contrast reversal.

The third experiment tested face and dog recognition for the effects of contrast reversal, that is, swapping the luminance values of a picture so that it looks like a photographic negative. Contrast reversal has been shown to have a very adverse effect on face recognition in both identification and matching tasks (Bruce & Langton, 1994; Galper, 1970; Johnston, Hill, & Carman, 1992; Kemp et al., 1990; Kemp, Pike, White, & Musselman, 1996; Phillips, 1972). However, for objects, in novices, the one published study found no effect of contrast reversal for greebles (Gauthier et al, 1998), and an unpublished conference presentation has also reported no effect for chairs (Subramaniam & Biederman, 1997).[6]

---

[5] The primary interest in Experiment 2 was in the expertise results. However, results for novices are also of interest, being relevant to the question of which behavioural tasks produce the most purely 'face-specific' effects. In both the inversion paradigm and the part-whole paradigm, there is a nonzero effect for objects, and the 'face-specificity' in novices is merely that the effect is much larger for faces than for objects. Results for the composite task, however, suggest the effect is present for faces and absent for objects. In Experiment 2 (N=23), novices showed no composite effect for upright dogs, despite a large number of trials and correspondingly small error bars (Figure 7). We also tested an additional independent group of young novices (N=24), who showed the same result (aligned = 85.45%, unaligned = 86.25%, $t < 1$). Combining data from all 47 novices we tested on dogs, a 95% confidence interval on the size of the aligned-unaligned difference was very tightly centred around zero (-1.44 – +1.47 % correct). Moreover, Gauthier and Tarr (2002) also found no composite effect in novices for greebles (Table 6). Overall, results argue that the composite effect dissociates faces and objects more strongly than does either inversion or part-whole. This suggests the composite effect provides a particularly pure measure of configural processing.

[6] Two additional studies claimed to have tested contrast reversal effects for objects in novices. Both used a same-different matching task, and report comparison of condition with one item normal contrast and one item reversed with a 'contrast matched' condition. One study showed no effect for objects (pigmented 'blobs', Nederhouser, Mangini, Biederman & Okada, 2003) and the

The contrast reversal effect for faces is not theoretically or empirically related to configural processing. Theoretically, contrast reversal tasks do not test integration of face parts into the whole. Empirically, inversion effects and contrast reversal effects do not interact (Bruce & Langton, 1994; Johnston et al., 1992; Kemp et al., 1990), and normal composite effects on the Young et al. task are still present for contrast reversed faces (Hole, George, & Dunsmore, 1999). These findings argue that contrast reversal effects and configural processing arise from different stages of visual processing.

However, given the findings that the difficulty of processing contrast reversed stimuli is much more severe for faces than other objects, it is a candidate 'face-specific effect' on which the effects of expertise can be assessed. Only one previous study has done so. Gauthier et al. (1998) found a contrast reversal effect on greeble identification of 10 ms in novices and 224 ms in experts, a significant increase with expertise. If this result is replicable, and extends to other object types, we would expect to see a large contrast reversal effect for dogs in our dog experts. Indeed, the potential value of shape-from-shading information would seem to be higher for labradors than greebles, given that labradors are natural objects and have gradual boundaries between parts. This suggests that, if anything, the effects of expertise should be larger in our experiment than in Gauthier et al.'s.

Our task required subjects to judge whether two dogs (or two faces), presented simultaneously for a brief period, were the same identity or different identity. Pairs were either both normal contrast, or both reversed contrast, with the primary interest being in whether the accuracy difference between these two conditions (i.e., the contrast reversal effect) was larger for dogs in experts than in novices. We also included pairs with one member normal contrast and the other reversed contrast, to encourage subjects to focus their judgements on matching identity rather than low-level visual appearance. Finally, we tested both upright and inverted orientations. This allowed a second look at whether inversion effects (particularly for normal-contrast dogs)

---

other showed a significant effect for greebles that was smaller than that for faces (Vuong, Peissig, Harrison & Tarr, 2005). In both cases, however, the 'contrast-matched' condition combined data from both-items-normal pairs and both-items-reversed pairs together, with data not reported for each condition separately. Thus, it is not possible to assess difficulty of increase with expertise, this time in simultaneous matching as opposed to long-term memory (Experiment 1).

Method

Subjects.

The experts (N=15) were as for Experiments 1 and 2. A group of young novices (N=20; age 18-30 years; 5 male; all Caucasian) also completed the experiment, for course credit (N=5) or $5 for the half-hour experiment. All subjects reported normal or corrected-to-normal vision.

Design.

For each subject group, stimulus class (dogs vs. faces), orientation (upright vs. inverted) and contrast condition (both original vs. both reversed vs. one original one reversed) were all manipulated within subjects. Class and orientation were blocked (e.g., a block of upright dogs) with block order counterbalanced across subjects. The dependent measure was accuracy to judge whether a simultaneous, briefly presented, pair of exemplars were of the same identity or different identity, regardless of whether the contrast was the same or not. As in Experiment 2, presentation times were longer for experts than for young novices.

Stimuli.

For each of the 60 faces and 60 dogs used in Experiment 1, a contrast reversed version (Figure 9) was created using the Inverse function in Photoshop. All 60 possible same identity pairs were created in the both original contrast format, and also in the both reversed contrast format. In the one-original-one-reversed format the reversed item could appear on either the left or the right, giving 120 of these pairs. For each subject, half of the same-identity pairs (i.e., 30

---

perception of contrast-reversed stimuli per se, as opposed to the difficulty of matching across two different contrast formats.

both original, 30 both reversed and 60 one-original-one-reversed**)** were chosen at random from the full set to be presented in each orientation condition.

For different identity pairs, each trial showed two different but similar exemplars (the pairings developed for the memory task in Experiment 1 were used; see Figures 1 and 2). To create the contrast reversal conditions the procedure used for the same identity pairs was repeated. In total there were 240 trials in a block (120 same-identity, 120 different-identity).

Procedure.

Apparatus was as for Experiment 2. Subjects were given 10 practice trials using chairs (stimuli courtesy of Bruno Rossion & Michael Tarr). Presentation time was always the same for all stimulus classes (faces and dogs, upright and inverted). For the young novices, on each trial, the stimulus pair was presented for 600 ms (or less if the subject responded earlier), again allowing subjects to look at each exemplar only 1-2 times. For the experts, presentation time was adjusted for each subject as described as in Experiment 2. The final time per trial ranged from 600-700 ms with longer durations generally given for older experts, and most tested at 650 ms. Subjects responded same-identity via one key, and different-identity via another. They were warned that they might feel they were guessing on some trials but were required to enter a response anyway, and were informed that 50% of trials would be same-identity. After response, there was a blank screen for at least 200-400 ms (longer if the subject responded before the stimulus had disappeared from the screen). On each trial, the position of each stimulus on the screen was jittered as in Experiment 2.

Results

The percentage correct, averaged over same- and different-identity trials, was calculated for the both original and both reversed conditions [7]. Results are shown in Figure 10.

[7] Accuracy in the one-original-one-reversed condition was consistently slightly poorer even than in the both reversed condition, across both subject groups, both stimulus classes, and both orientations. This suggests, unsurprisingly, that comparison of identity across two different contrasts is more difficult than comparison of identity when contrast is held constant.

Contrast reversal and inversion effects in young adult novices.

For young novices, effects of contrast reversal and inversion were independent, with no 2-way orientation x contrast condition interaction, $F(1,19)= 1.45$, $MSE= 25.33$, $p>.2$, and no 3-way orientation x contrast x class interaction, $F(1, 19) = 1.66$, $MSE = 16.74$, $p > .2$. The contrast reversal effect was significantly greater for faces than for dogs, $F(1, 19) = 39.97$, $MSE = 25.86$, $p < .001$ (class x contrast interaction). For faces (Figure 10a), contrast reversal impaired identity-matching both upright, $t(19)=10.47$, $p<.001$, and inverted, $t(19)=7.49$, $p<.001$. The small contrast reversal effect for dogs (Figure 10b) was also significant for both upright, $t(19)=2.91$, $p<.01$, and inverted, $t(19)=4.07$, $p<.001$. Inversion also had a larger effect for faces than for dogs, $F(1, 19) = 8.15$, $MSE = 41.74$, $p < .01$ (class x orientation interaction). The inversion effect for faces was significant for both normal contrast pairs, $t(19)=5.88$, $p <.001$, and reversed contrast pairs, $t(19)=3.41$, $p <.01$. For dogs, there was no inversion effect in either case, $t < 1$ normal contrast, $t < 1$ reversed contrast.

These results are consistent with previous findings in novices. That is, for faces, both contrast reversal and inversion each substantially impaired recognition, although the two effects did not interact, suggesting that they arise from different stages of processing. For dogs, however, both contrast reversal and inversion had very small effects (indeed, nonsignificant in the case of inversion).

Contrast reversal and inversion effects in experts.

For experts, contrast reversal and inversion interacted rather than showing independence, $F(1,14)=9.26$, $MSE=29.69$, $p <.01$ (2-way contrast x orientation). The origin of this interaction was primarily that, for faces (Figure 10c), the effect of contrast reversal was large for upright, $t(14)=6.70$, $p <.001$, but small and nonsignificant for inverted, $t < 1$. This finding was different from that in our young novice group, and also from that in previous studies of faces, but is probably not theoretically interesting, as it could simply reflect the older subjects' performance approaching floor for inverted faces. The more interesting question was what happened for dogs

(Figure 10d). Here, experts showed <u>no</u> contrast reversal effect, either upright, t <1, or inverted, t <1; indeed, the slight trend was in the opposite-to-expected direction for a contrast reversal effect (reversed better than original). The contrast reversal effect in the expert group was then significantly larger for faces than for dogs, $F(1,14)=11.93$, $MSE=52.83$, $p<.01$.

Turning to orientation, the expert group showed a large inversion effect for normal contrast faces, $t(14)=3.93$, $p<.01$. There was no inversion effect for contrast reversed faces, $t < 1$, probably again due to the floor effect. Critically, for dogs, there was no effect of orientation for either contrast condition, ts<1.

These results for experts show no evidence of face-like processing. Dog experts showed no contrast reversal effects or inversion effects for dogs.

<u>Statistics comparing patterns in experts and novices</u>.

A 4-way mixed ANOVA directly compared experts and novices as the between subjects factor, with class (faces vs. dogs), contrast condition (both normal vs. both reversed) and orientation (upright vs. inverted) as within-subjects factors. In testing the expertise hypothesis, the critical question is whether there was any indication that dog experts looking at dogs were more strongly affected by either contrast reversal or inversion than novices. Obviously there was no indication in the means (Figure 10) of any such effects, and the ANOVA confirmed no 4-way or 3-way interactions involving experience and either contrast reversal or inversion (largest $F=2.70$, $p=.11$).

The ANOVA did reveal two significant interactions involving expertise. The first reflected the fact that experts had better overall performance for their objects-of-expertise than novices, despite worse performance on faces, regardless of contrast or orientation, $F(1,33)=24.08$, $p<.001$ (2-way class x experience). As well as reflecting psychophysical evidence of expertise for the upright normal contrast condition, this is an interesting finding in that it is relevant to the <u>type</u> of expertise our experts had gained by lifetime exposure to dogs: whatever this is, it could be applied to dogs in <u>all</u> formats even highly unfamiliar ones (inverted and contrast reversed). The finding is also inconsistent with the expertise hypothesis, which explains large inversion and contrast reversal effects for faces based on experience being obtained in the upright and normal contrast form. The second was a 2-way contrast x experience interaction, $\underline{F}(1,33) = 22.22$, $\underline{p} < .001$. Again, this did not follow the direction predicted by the expertise hypothesis. Experts showed <u>smaller</u> contrast reversal effects than novices, for <u>both</u> faces and dogs, and in both orientation conditions (Figure 10).

<u>Correlations with experience for experts</u>

Within the expert group, there was no correlation between the size of the contrast reversal effect for upright dogs (normal contrast minus reversed contrast) and either years of experience, $\underline{r} = -.16$, $\underline{p} > .5$ (Figure 11), or lifetime number of dogs seen, $\underline{r} = -.10$, $\underline{p} > .7$.

<u>Discussion</u>

In terms of the effect of contrast reversal, Experiment 3 found no evidence of face-like processing for dogs in dog experts. Despite confirming a large contrast reversal effect for faces, experts showed no sensitivity to contrast reversal for dogs under the same presentation conditions. Only one previous study (Gauthier et al., 1998) has examined contrast reversal for objects-of-expertise, finding that expertise increased this effect for greebles in experiment trained experts relative to novices. This finding seems surprising in the context of our present results. If anything, the expertise effect on sensitivity to contrast reversal could have been expected to be larger for dogs than for greebles, given that dogs are a natural object class for which shape-from-shading is potentially a very useful cue, and that our experts had far more experience than did Gauthier et al.'s. We also note that Gauthier et al.'s 'novices' had in fact received some degree of training, sufficient to allow naming of the individual target greebles. Thus, if experience is, indeed, the key factor in predicting contrast reversal effects, we might have expected their study to have shown a rather larger contrast reversal effect in the novice group than the tiny 10 ms difference they obtained (out of a baseline of approximately 1580 ms).

Turning to orientation, results of Experiment 3 found that experts showed a large inversion effect for faces, but no inversion effect for dogs. This replicates the finding of Experiment 1, using simultaneous matching task rather than long-term memory. The data also speak to a possible measurement issue regarding the size of the inversion effect across stimulus classes. In Experiment 1, readers may have been concerned about the fact that experts' performance was not equal for upright faces and upright dogs (Figure 3), and thus the inversion effect for the two stimulus classes was in some sense out of different 'baselines' (instead, in that experiment, we had pilot-tested stimuli to equate faces and dogs in the inverted orientation in novices, as this guaranteed room to reveal better performance for upright dogs in experts). In Experiment 3, however, upright faces and upright dogs were equal in experts (Figure 10). Thus, regardless of whether faces and dogs were equal on upright performance (Experiment 3) or inverted performance (Experiment 1), experts showed a much bigger inversion effect for faces than dogs.

Finally, Experiment 3 indicated that the type of expertise gained by lifetime exposure to dogs could be applied to dogs even in highly unfamiliar formats including inverted and contrast reversed (Xu et al., 2005, report a similar result in which car expertise improved matching of both upright and inverted cars). We have no direct evidence as to the nature of processing supporting this expertise; however, we know it cannot be configural, given the lack of composite effect for dogs in dog experts in Experiment 2. A reasonable hypothesis is that expertise with dogs improved part-based processing. Division of the stimulus at contour boundaries is commonly proposed to form the input to part-based object recognition (e.g., Biederman, 1987), and inversion and contrast reversal leave this contour boundary information intact.

General Discussion

In dog experts, our results showed no evidence of face-like processing for objects-of-expertise on three behavioural tasks. This included only weak and nonsignificant sensitivity to inversion in recognition memory (Experiment 1) and no sensitivity to inversion in simultaneous matching (Experiment 3), no Young et al. (1987) composite effect for dogs even when upright (Experiment 2), and no sensitivity to contrast reversal (Experiment 3). Further, experts showed no increases relative to novices, in inversion, composite, or contrast reversal effects. We also emphasise that it was not the case that we found no significant effects in the expert group. Like novices, experts showed all the expected findings for faces, using the same procedures as were used for dogs. These included large and significant inversion effects, a significant composite effect for upright faces, and large and significant contrast reversal effects.

Our results were obtained despite the processing and experience demands for dogs being as similar as possible to those for faces, conditions which should have given the best possible chance for data supporting the expertise hypothesis to emerge. Specifically: we chose a stimulus class where experts look at individual identity; we used tasks requiring processing of individual identity; we selected experts with many years of experience (indeed, many had the same number of years' experience with dogs as the young novices had with faces); and we tested a natural object class, with genetic variability, and for which shape-from-shading information is potentially useful. Our results therefore argue strongly against the expertise hypothesis, and favour domain-specificity of 'special' processing for faces.

A final finding was that expertise improved processing of dogs (relative to faces) in all forms, not just the canonical (Experiment 3). In the canonical upright normal contrast form, novices were poorer with dogs than faces, but experts showed equal performance for the two stimulus classes. In noncanonical forms (inverted and/or contrast reversed), novices were approximately equal with dogs and faces, but experts were better with dogs than faces (see Figure 10). Together with the lack of configural processing (Experiment 2), this finding suggests that the effect of perceptual expertise in our dog experts was to improve part-based processing for dogs.

Evidence of Expertise: Revisited

We have presented considerable evidence that our experts were, indeed, expert with dogs. In addition to the number of years experience, estimated number of dogs seen, and anecdotal evidence of expertise described in the introduction (see 'Evidence of expertise in our subjects' section), we provided two experimental tests. Both of these were conducted with upright, normal contrast dogs, shown complete and intact. It is under these circumstances that evidence of expertise would clearly be expected. Most powerfully (i.e., 60 trials per condition per subject), in

Experiment 3 experts were as good at matching identity for dogs as for faces, despite novices showing substantially poorer performance with dogs than faces. In Experiment 1, there was also a trend towards better recognition memory for dogs in experts than in age-sex-education matched controls, which was significant on a 1-tailed test for the experts with more than 10 years of experience (N=12) versus their controls. We are satisfied with this result given the high variability of recognition memory data resulting from the much reduced number of trials that can be employed in this type of paradigm (15 per condition in our case). The size of the expertise advantage (11% in our 12 most experienced experts; Table 1) was also similar to that found in the only previous study to examine recognition memory in experts and age matched novices (11% for handwriting; Bruyer & Crispeels, 1992).

Despite this evidence of expertise in our subjects, there are a number of places throughout the experiments where at first glance, it might be thought that a basic expertise effect (i.e., experts better than controls with upright dogs) would be expected, and such a result was not obtained. We now discuss each of these cases in turn, explaining why advantages for experts would not necessarily be predicted.

In Experiment 1 (recognition memory task), experts showed an advantage with upright dogs compared to age matched controls, but no advantage compared to young novices. It has been suggested to us that the comparison of experts and young novices could be valid, despite the age difference, because experts and young novices showed equal performance on faces. However, it needs to be kept in mind that performance on a memory task is a combination of perception of dogs, which indeed ought to be better in experts than novices, and memory ability, which is probably poorer in the experts than the young novices given the substantial mean age difference; thus, drawing conclusions about dog perception by comparing these two groups is not valid, even if the face performance is matched. An alternative interpretation of our data is that memory for faces has some special protection against the usual decline in memory with age. Much of the usual age-related decline in memory performance is due to lack of spontaneous use of appropriate learning and retrieval strategies (e.g., allocating insufficient attention), rather than a real inability to remember once appropriate strategies are used (see Hertzog & Hultsch, 2000). Faces may then have special protection because they remain attention-attracting, socially salient stimuli to older subjects, automatically encouraging good memory strategies. Empirically, our data from novices are consistent with this idea. Looking at the two novice groups in Figure 3, memory for upright faces remains stable with age, while the general pattern is that memory for all other classes (inverted faces, upright dogs, inverted dogs) declines as expected.

In the next result to be considered, Experiment 2 (composite task) found that experts were not noticeably better than young novice controls with upright dogs in either the aligned or unaligned conditions. Of course, this comparison is invalid simply because of the age difference between the two groups and the fact that we gave the older experts a slightly easier version of the procedure (longer presentation duration). More importantly, even had we included an age-matched control group given identical procedures to the experts, the comparison of experts and novices would still not be useful in assessing behavioural expertise. This is because there is no clear prediction regarding the direction of an expertise effect: experts are familiar with whole dogs, but the task in this case is to match half dogs. It is not clear whether perceptual expertise would help or hinder performance relative to age-matched controls in this task (i.e., experts could find half dogs more disturbing than novices).

In Experiment 3 (identity matching), readers should note that, again, the overall comparison of expert and novice accuracy is not valid because experts were older and given longer presentation durations. What is valid in this experiment is the comparison of faces and dogs within each group because the same subjects and procedures were used in both cases. As described earlier, results showed powerful evidence of behavioural expertise (dogs as good as faces in experts, and $p<.001$ on the interaction of stimulus class and expertise level).

Expertise hypothesis and configural/holistic processing: The standard tasks.

A major aim of this article has been not just to present our own new data, but also to provide clear summaries of previous findings. The review component has been important because it points out that our results are not as surprising as we suspect many readers will find them. There has been a widespread assumption in the literature that there is much evidence from behavioural studies to support the expertise hypothesis. However, much of this evidence has unfortunately come from miscitation, or, in a few cases, from selective reading of the data (e.g., focussing on a close-to-significant expertise effect in one condition, while ignoring the presence of null or reverse effects in other equally relevant conditions). We have included tables containing the actual data from the original studies to demonstrate that the results did not contain the effects of expertise often claimed.

The primary theoretical question on which most of this article has focussed is whether expertise with objects induces configural/holistic processing, the main style of computation known to be specific to faces in novices. In assessing this, results of three standard paradigms – inversion, part-whole, and composite – are of relevance, either indirectly or directly.

Disproportionate inversion effects have been theoretically associated with configural processing by many authors, although note that this association is only indirect, in that large inversion effects have been shown to originate from configural processing for faces, but there is no logical guarantee that a similar effect for objects must come from this source. Despite this theoretical limitation, inversion studies have produced the only case of strong supportive evidence for the expertise hypothesis, namely Diamond and Carey's (1986, Experiment 3) classic finding of inversion effects for dogs in dog experts that were as large as those for faces. Unfortunately, however, our review shows that this result does not replicate with a range of other object classes (Table 3). Even more seriously, our own study shows that the result does not replicate even when the stimuli and experts are very similar to those of Diamond and Carey: that is, like their study, we used dogs shown in side-on views, and ensured our experts had specific expertise with the breed. We have suggested that Diamond and Carey's unusual result could be attributable to their subjects having been pre-experimentally familiar, in the upright orientation, with some of the dog images used as stimuli. Overall, the evidence shows that inversion effects change little, if at all, with expertise, and certainly do not become face-like.

Turning to more direct measures of configural/holistic processing, both the part-whole and composite tasks are widely accepted as standard in the literature, and both of these have now been tested for a range of object classes (manmade and natural) and expert types (experiment-trained and real world). In the part-whole task, previous studies had already shown no effect of expertise: in comparison to the large part-whole effect for faces, both experts and novices show a small effect for objects that is no bigger in experts than in novices (Table 5). Thus, the part-whole results fail to support the expertise hypothesis. In the composite paradigm, prior to the present study, results were available from only two expertise studies, both testing manmade objects in experiment trained experts. Their results, showing no composite effect in greeble experts (Table 6), agree with our finding of no composite effect in dogs. Thus, again, the composite task results fail to support the expertise hypothesis for both natural and artificial objects, and for real world as well as experiment-trained experts. The conclusion from standard paradigms is clearly that configural/holistic processing does not occur for objects-of-expertise.

Expertise hypothesis and configural/holistic processing: Other tasks.

Before leaving configural processing, however, we are obliged to consider results from other, less common, paradigms that are (or have been claimed to be) relevant to assessing configural processing style. First is Tanaka and Sengco's (1997) whole-vs-configurally-transformed-whole modification of the original part-whole paradigm. Here, for faces, memory for a part (Mary's mouth) is worse in the context of a configurally altered version of the original face (Mary's mouth in Mary's face with the eyes shifted further apart) than in the unaltered version (Mary's mouth in Mary's face). We agree with other authors (Tanaka & Sengco, 1997; Gauthier & Tarr, 2002) that this shows strong evidence of configural/holistic processing; for faces, it proves that the usual whole-over-part advantage does not come merely from the presence of the extra context provided by having more features in the whole condition than in a part-alone condition, but instead depends on reinstating the particular configuration of the studied features. This modified paradigm has been tested in three expertise studies, all using greebles. Results of the first study were quite suggestive of an expertise effect: Gauthier and Tarr (1997) found a significant difference between part-in-original-whole and part-in-configurally-transformed-whole for greeble experts on reaction time (although there was no effect on accuracy), with a close-to-significant interaction involving expert-vs-novice status. However, the second and third studies failed to confirm the finding. In Gauthier et al. (1998) only one of the three greeble parts tested showed a close-to-significant effect in experts, and the other two parts showed trends in the reverse-to-predicted direction. Averaged across all parts, the mean effect was 0 percentage points for novices and 0 percentage points for experts (cf. 5-11 points for faces in Tanaka & Sengco, 1997). Gauthier and Tarr (2002) again tested all three parts, reporting both accuracy and reaction time. In experts, there was one significant effect in the predicted direction (reaction time for the 'quiff'), but there was also another significant effect in the reverse-to-predicted direction (accuracy for the 'dunth'). Averaged over all parts, the size of the effect in d' accuracy was 0.69 in session 1 (novices) and 0.64 in session 5 (experts); that is, the effect did not increase with expertise.

A second technique is that introduced by Gauthier, Curran, Curby and Collins (2003), using car experts. Here, we disagree with the authors that the method assesses face-like configural/holistic processing at all. Their method bore some resemblance to the Young et al. (1987) composite paradigm. However, a very important difference is that they did not make the usual comparison between aligned and unaligned stimuli. Such a comparison ensures that response competition from the nontarget half is matched for the two conditions, meaning that worse performance in the aligned condition can only be attributed to integration of parts into a new whole at the perceptual level (rather than being a response/decision effect). Instead, they used a much less stringent definition of 'holistic' processing as merely the inability to ignore a notionally irrelevant half (specifically, ‘obligatory processing of all features of an object, when subjects are instructed to attend selectively to one feature while ignoring others’, p. 1). In a sequential same-different task, they contrasted two conditions in which the response suggested by matching the nontarget half was either consistent with the response suggested by the target half (e.g., the nontarget top halves were the same when the target bottom halves were the same) or inconsistent with the response suggested by the target half (e.g., the nontarget top halves were different when the target bottom halves were the same). They took the difference between these conditions as their measure of holistic processing, finding that the difference between ‘holistic processing’ for normal cars versus transformed cars (the top half was inverted) increased with expertise. In some sense, of course, there is nothing wrong with the Gauthier et al. (2003) definition of 'holistic' processing given the many ways this term has been used in psychology; however, it does not assess the same type of perceptual integration as is known to occur for faces. Indeed, under their definition, practically any two things one cared to test would be expected to be processed 'holistically', given that our attentional system is rarely able to completely tune out irrelevant stimuli or components of the stimulus. For example, the Stroop effect (i.e., the difficulty of ignoring the word red when trying to name the colour of the ink it is printed in as blue) would be interpreted as showing that colour and word identity are processed together holistically, despite the fact that colour and word form/names are clearly not integrated at a perceptual level (e.g., they are processed by very different brain regions). Thus, the effect of expertise revealed in the Gauthier et al. study is, to our minds, not an expertise effect on face-like holistic processing; instead, it merely shows that response competition from the to-be-ignored half is stronger when subjects are more familiar with the cars. This seems unsurprising, in that greater familiarity with cars should mean they are processed more quickly, presumably making irrelevant halves harder to ignore.

A third technique has recently been used by Busey and Vanderkolk (2005). They studied fingerprint experts, showing that performance in an X-AB matching task (study X, then choose A or B as the old item) was better for experts than novices, with the expertise benefit most obvious when the stimuli were presented in noise. Their suggestion that this expertise effect reflected configural processing was based on comparing a full-image condition with a half-image condition (in which several areas of the stimulus were masked), and fitting a multinomial choice model. The results of this showed that the estimated probability of accepting the target in the full-image case and the half-image case differed, in a direction that they argued showed that experts acquired more information from a given half of the image when the other half is present than when the other half is absent. We agree with Busey and Vandervolk that a demonstration of more-than-additive effects of a second half would provide good evidence of configural processing. However, acceptance of their conclusion requires acceptance of the validity of their model. This seems in some doubt because, while the result just described was obtained for experts when fingerprints were in noise, a peculiar opposite-direction result was found in other conditions (including experts without noise), apparently arguing that subjects acquire less information from each half when the full image is presented than when each half is presented alone. It is not clear that this result should be logically possible: surely the worst possible outcome even in novices should be independence of the two halves, not underadditivity. Further, if it is true that experts have access to configural information that can help them 'see through the noise', it is not obvious why they would not also use this configural information to enhance their performance without noise. These results suggest to us that something might be wrong with the model.

What might this be? Busey and Vanderkolk (2005) note that their results, rather than reflecting configural processing, could be explained by experts and novices having different nonlinear transfer functions between image luminance and the visual system's response. In particular, they suggest that, over the course of a career involving examining thousands of low-contrast fingerprints, experts might have altered their sensitivity at low brightness levels compared to novices, and this would affect the validity of the model's comparison of experts and novices.

Whether or not this is the real source of the problem in their model, the argument that Busey and Vanderkolk's expertise results reflect configural processing is uncompelling. We particularly note that faces were not tested in their additivity experiment. To have confidence in their approach and underlying model, one would wish to know that it produces the expected results for faces (i.e., overadditivity for upright faces, and independent processing of halves for inverted faces; see Yovel, Paller & Levy, 2005).

Overall, of the three additional tasks reviewed in this section, none supports the expertise hypothesis. One measures face-like configural/holistic processing (the Tanaka & Sengco, 1997, procedure), but has shown no expertise effects for greebles. The other two have shown clear expertise effects, but do not (Gauthier et al., 2003) or quite possibly do not (Busey & Vanderkolk, 2005) assess the same type of configural/holistic processing as used for faces.

<u>The expertise hypothesis and contrast reversal.</u>

As noted earlier, sensitivity to contrast reversal has no obvious relationship to configural processing, but is of potential relevance to evaluating the expertise hypothesis in that a strong sensitivity to contrast reversal appears to be 'special' to faces in novices. In terms of expertise effects, results are now available from two studies. Gauthier et al. (1998) found sensitivity to contrast reversal increased with expertise for experiment-trained greeble experts. However, our own study failed to replicate this with real world dog experts, under circumstances where, if anything, we would expect a greater chance of finding contrast reversal effects. The reason for the difference in results is currently unclear. Resolution awaits testing of additional object classes and expert types.

The expertise hypothesis and other behavioural and neural phenomena

The primary purpose of this article has been to examine behavioural findings testing configural processing. It has also been suggested that objects-of-expertise may demonstrate similar processing to faces in several other ways, both behavioural and neural. We finish with a few brief comments on this literature, noting that although several research domains show results that could be taken as supporting the expertise hypothesis, none do so unambiguously.

In terms of cortical specialisation, we reviewed in the introduction findings of a double dissociation between faces and objects-of-expertise in neuropsychological patients; this argues against the expertise hypothesis. We also considered fMRI findings, noting that most fMRI results show a small increase in activation in the FFA for objects-of-expertise as compared to the same objects in novices. This could be taken as supporting the expertise hypothesis, but it could alternatively reflect attentional effects or limits on spatial resolution of fMRI (see introduction).

In terms of EEG data, faces produce an inversion effect in which the N170 response is larger and delayed for inverted faces as compared to upright faces (Rossion, Gauthier, Tarr, Despland, et al., 2000). Objects have been shown to produce the same effect in experts and not novices (fingerprints, Busey & Vandervolk, 2005; other studies can be criticised on methodological grounds such as sensor location, see McKone & Kanwisher, 2005). This could be taken to support the expertise hypothesis. However, the spatial resolution of EEGs is very poor and the apparently similar effects for faces and objects-of-expertise could be coming from different locations within the fusiform region (or even from other cortical areas).

A final finding is the downward shift in categorisation level in behavioural studies of expertise. Experts are as fast to identify their objects-of-expertise at the subordinate level (e.g., Volkswagon Beetle) as at the basic level (car), although the basic level is preferred in novices (e.g., Tanaka & Taylor, 1991). Tanaka (2001) found that individual identification of faces required only the same processing time as did basic-level categorisation (i.e., categorisation as a face), suggesting an interesting similarity between faces and objects-of-expertise. However, Grill-Spector and Kanwisher (2005, p. 158) failed to replicate Tanaka's result (instead, individual-level decisions for faces required longer than basic level decisions; also see Robbins & McKone, 2003, p. 88), and so the status of the downward shift comparison currently remains unclear.

Conclusion

In agreement with neuropsychological evidence of a double dissociation in cortical location between faces and objects-of-expertise, behavioural studies investigating style of computational processing do not support the predictions of the expertise hypothesis. Investigations of configural/holistic processing are, instead, supportive of true domain-specificity for faces. The focus of future research should be on why this arises; that is, the extent to which 'special' processing for faces is based on innate factors, and the extent to which it is driven by exposure to faces at particular periods of development, such as childhood or early infancy.

References.

Anglin, A. D., Hser, Y., & Chou, C. (1993). Reliability and validity of retrospective behavioural self-report by narcotics addicts. *Evaluation Review, 17*(1), 91-108.

Behrmann, M., Marotta, J., Gauthier, I., Tarr, M. J., & McKeeff, T. J. (2005). Behavioral change and its neural correlates in visual agnosia after expertise training. *Journal of Cognitive Neuroscience, 17*, 554-567.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception, 23*(7), 803-822.

Bruyer, R., & Crispeels, G. (1992). Expertise in person recognition. *Bulletin of the Psychonomic Society, 30*(6), 501-504.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45*(4), 431-448.

Carey, S. (1992). Becoming a face expert. *Philosophical Transcripts of the Royal Society London: B 335*, 95-102; discussion 102-103.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1*(2/3), 253-274.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using

Macintosh computers. *Behavior Research Methods, Instruments and Computers, 25*, 257-271.
Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica, 73*(3), 225-243.
de Haan, M., Humphreys, K., & Johnson, M. H. (2002). Developing a brain specialized for face perception: A converging methods approach. *Developmental Psychobiology, 40*(3), 200-212.
De Renzi, E. (1986). Current issues on prosopagnosia. In H. Ellis, M. A. Jeeves, F. Newcombe & A. Young (Eds.), *Aspects of face processing* (Vol. 28, pp. 243-252.). Dordrecht: Martinus Nijhoff Publishers.
Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117.
Dixon, M. J., Desmarais, G., Gojmerac, C., Schweizer, T. A., & Bub, D. N. (2002). The role of premorbid expertise on object identification in a patient with category-specific visual agnosia. *Cognitive Neuropsychology, 19*(5), 401-419.
Donnelly, N., & Davidoff, J. (1999). The mental representations of faces and houses: Issues concerning parts and wholes. *Visual Cognition, 6*(3-4), 319-343.
Duchaine, B. C., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron, 43*(4), 469-473.
Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioural Brain Research, 76*(1-2), 181-189.
Fastenau, P. S., Denburg, N. L., & Abeles, N. (1996). Age differences in retrieval: Further support for the resource-reduction hypothesis. *Psychology and Aging, 11*(1), 140-146.
Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19*(4), 207-208.
Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*(4), 428-432.
Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191-197.
Gauthier, I., & Tarr, M. J. (1997). Becoming a 'greeble' expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.
Gauthier, I., & Tarr, M. J. (2002). Unravelling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance, 28*(2), 431-446.
Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. *Vision Research, 38*(15-16), 2401-2428.
Grill-Spector, K. & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*, 152-160.
Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555-562.

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

Hertzog, C. & Hultsch, D. (2000). Metacognition in adulthood and old age. In F. I. M. Carik & T. A. Salthouse (Eds.), *The handbook of ageing and cognition* (2nd ed., pp. 417-466). Mahwah, NJ: Erlbaum.

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

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

Johnston, A., Hill, H., & Carman, N. (1992). Recognizing Faces - Effects of Lighting Direction, Inversion, and Brightness Reversal. *Perception, 21*(3), 365-375.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*(8), 759-763.

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

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception, 19*(4), 531-543.

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*(1), 37-52.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 410*, 890.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6*(10), 1108-1112.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science, 15*(11), 762-768.

Leder, H., & Carbon, C.-C., (2005). When context hinders! Learn-test compatibility in face recognition. *Quarterly Journal of Experimental Psychology, 58A*, 235-250.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*(5), 605-610.

McKone, E., & Kanwisher, N. (2005). Does the human brain process objects of expertise like faces? A review of the evidence. In G. Rizzolatti (Ed.), *From monkey brain to human brain.* Cambridge, Mass.: MIT Press.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: a face-specific disorder. *Quarterly Journal of Experimental Psychology: A, 46*(1), 1-10.

Meadows, J. C. (1974). Anatomical basis of prosopagnosia. *Journal of Neurology Neurosurgery and Psychiatry, 37*(5), 489-501.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review, 98*(2), 164-181.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Psychology, 9*(5), 555-604.
Nederhouser, M., Mangini, M. C., Biederman, I., & Okada, K. (2003). Invariance to contrast inversion when matching objects with face-like surface structure and pigmentation [Abstract]. *Journal of Vision, 3*, 93a.
Nicholas, A. K. (1983). *The book of the labrador retriever*: TFH publications, Inc. Ltd.
Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296*(5571), 1321-1323.
Peterson, M., & Rhodes, G. (2003). Analytic and holistic processing- The view through different lenses. In M. Peterson & G. Rhodes (Eds.), *Perception of Faces Objects and Scenes: Analytic and Holistic Processes.* (Vol. 2, pp. 3-20.). Oxford: Oxford University Press.
Phillips, R. J. (1972). Why are faces hard to recognize in photographic negative? *Perception and Psychophysics, 12*(5), 425-426.
Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275-&.
Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience, 16*(2), 189-203.
Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition, 88*(1), 79-107.
Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science, 13*(3), 250-257.
Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linnotte, S., et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuro Report, 11*, 69-74.
Schwarzlose, R. F., Baker, C. I. & Kanwisher, K. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience, 25*(47), 11055-11059.
Subramaniam, S., & Biederman, I. (1997). Does contrast reversal affect object recognition? [Abstract]. *Investigative Ophthalmology and Visual Science, 38, 998.*
Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130*, 534-543.
Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12*(1), 43-47.
Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: A, 46*(2), 225-245.
Tanaka, J. W., & Gauthier, I. (1997). Expertise in object and face recognition. In R. L. Goldstone, D. L. Medin & P. G. Schyns (Eds.), *Mechanisms of Perceptual Learning* (Vol. 36, pp. 83-125.). San Diego, CA.: Academic Press.
Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25*(5), 583-592.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(3), 457-482.
Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. Peterson & G. Rhodes (Eds.), *Perception of Faces Objects and Scenes: Analytic and Holistic Processes* . (Vol. 2., pp. 177-211.). Oxford: Oxford University Press.
Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17*(1-3), 257-279.
Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*(Pt 4), 471-491.
Vuong, Q. C., Peissig, J. J., Harrison, M. C., Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and greebles. *Vision Research*, 45, 1213-1223.
Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*(1), 59-85.
Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology, 79*(3), 1574-1578.
Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*.
Xu, Y., Liu, J., & Kanwisher, N. (2005). The M170 is selective for faces, not for expertise. *Neuropsychologia, 43*(4), 588-597.
Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141-145.
Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747-759.
Yovel, G., Paller, K. A., & Levy, J. (2005). A whole face is more than the sum of its halves: Interactive processing in face perception. *Visual Cognition*.

Acknowledgements

Research supported by Australian Research Council grants DP0208630 and DP0450636 to EM. We are very grateful to the dog experts for agreeing to participate, and for their assistance in contacting additional experts. We thank Mary Conheady of the Victoria Labrador Retrievers Association for her assistance in finding pictures of dogs, and Jacqui Brewer for assistance in locating and testing age-sex-education matched controls in Experiment 1. Address correspondence to: Elinor McKone, School of Psychology, Australian National University, ACT 0200, Australia. Email: Elinor.McKone@anu.edu.au, or Rachel Robbins who is now at Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, L8S 4K1, CANADA, Email: Dr.R.Robbins@gmail.com

Table 1. Mean (and SEM) of accuracy on recognition memory task (% correct), for the experts with ≥ 10 years experience (N=12), and novices matched to experts one-to-one for age, sex and education level (N=12).

	Upright faces	Upright dogs
Experts	81.7	74.4
	(4.2)	(4.6)
Age-matched novices	83.3	63.3
	(4.0)	(3.7)

Note: chance = 50%

Table 2. Mean accuracy on identity-matching task (% correct) for all experts (N=15) and young adult novices (N=20). The SEM reported in brackets is appropriate for the within-subjects comparison of faces and dogs (i.e., SEM of faces – dogs difference score).

	Experts	Young Novices
Upright faces	75.9	85.2
	(4.1)	(2.2)
Upright dogs	76.4	74.3
	(4.1)	(2.2)

Notes: 1. chance = 50%

2. Direct comparison of Experts and Young novices is invalid due to age and presentation time differences (see Experiment 3). Comparison of faces versus dogs within each group is valid.

Table 3. Results of the previous studies that have examined inversion effects for objects-of-expertise, showing the inversion decrement (upright – inverted). Studies reported various measures, including percent correct (%), d', and reaction time (ms). For novices and experts, the significance or otherwise of each inversion effect is indicated; a separate column indicates whether the increase in the size of the inversion effect from novices to experts was significant. Results are also provided for faces (based on data from the experts group, or averaged across experts and novices where there were no differences).

	task	Novices	Experts	sig of expertise increase	Faces
Dogs (Diamond & Carey, 1986 - Experiment 3)	long-term memory	5% ns	22% *	*	20% *
Handwriting (Bruyer & Crispeels, 1992)	long-term memory	5% ns	9% *	ns (F<1)	20% *
Greebles (Rossion et al., 2002)	sequential matching	25 ms –	46 ms –	*	75 ms –
Cars (Gauthier et al., 2000)	sequential matching	d' = .57 *	d' = .84 *	–	–
Birds (Gauthier et al., 2000)	sequential matching	d' = .05 ns	d' = .30 ns	–	–
Cars (Xu et al, 2005)	sequential matching	8% – d' = .44 –	8% – d' = .87 –	ns *	– –
Fingerprints (Busey & Vanderkolk, 2005)	face/print classification	– ns	– ns	ns	– *

Notes: * = p<.05; ns = p>.05; reverse = trend in opposite-to-predicted direction; – = not tested or not reported.

Table 4. Reaction times (in ms) for Experiment 1: Recognition memory task. The SEM shown is appropriate for the within-subjects test of the significance of the inversion effect (i.e., SEM of inverted – upright difference score)

	Faces		Dogs	
	upright	inverted	upright	inverted
Young novices	3027	4521	4177	4487
	(253)	(253)	(261)	(261)
Age-matched novices	4046	4943	5229	5588
	(271)	(271)	(328)	(328)
Experts	3191	4416	4419	4997
	(316)	(316)	(457)	(457)

Note: Results shown are for all trials, not only correct ones. The results for correct-only trials were essentially the same, but less reliable given the smaller number of trials.

Table 5. Results of previous studies using the Tanaka & Farah (1993) part-whole paradigm for objects-of-expertise, showing size of the whole – part difference, averaged over all parts tested. All stimuli were upright. Formatting as for Table 3.

	Novices	Experts	sig of expertise increase	Faces
Dog faces (Tanaka et al., 1996)	2% ns	8% ns	ns	20% *
Cars (Tanaka et al., 1996)	8% –	6% –	reverse	18% *
Biological cells (Tanaka et al., 1996)	16% *	10% *	reverse	26% *
Greebles (Gauthier & Tarr, 1997)	5% ns	11% *	ns	–
Greebles (Gauthier et al., 1998)	7% –	0% –	reverse	–
Greebles (Gauthier & Tarr, 2002)	d' = 0.75 –	d' = 0 .68 –	reverse	–

Notes: * = p<.05; ns = p>.05; reverse = trend in opposite-to-predicted direction; – = not tested or not reported.

Table 6. Results of previous studies of the Young et al. (1987) composite paradigm for objects-of-expertise, showing the aligned – unaligned difference (for reaction times) or unaligned – aligned for accuracy (in both cases, a positive number corresponds to the direction for a positive composite effect, i.e., aligned should be the more difficult condition). All stimuli were upright. Formatting as in Table 3.

	Novices	Experts	sig of expertise increase	Faces
Greebles, same-family halves (Gauthier et al., 1998)	–	115 ms ns 0%	– –	–
Greebles, different-family halves (Gauthier et al., 1998)	–	–37 ms reverse –3% reverse	– –	–
Greebles, same-family halves (Gauthier & Tarr, 2002)	–42 ms –	12 ms –	– [a]	–

Notes: * = p<.05; ns = p>.05; reverse = trend in opposite-to-predicted direction; – = not tested or not reported.

[a] Across 5 sessions (we show only session 1 = novices, session 5 = experts), there was a close-to-significant interaction between session and aligned vs. unaligned. However, this did not reflect an increase with expertise: the composite effect started close to zero, strangely became more negative in sessions 2-4, then returned to close to zero. Also note the 12 ms composite effect in experts was in the context of 35 ms SEMs for the aligned and unaligned conditions.

Table 7. Experiment 2 (composite task): Accuracy for different-identity trials (% correct). Numbers in brackets are the within-subjects equivalent of the SEM for making the comparison between aligned and unaligned conditions.

	Faces		Dogs	
	Aligned	Unaligned	Aligned	Unaligned
Experts				
Upright	69.44	64.22	57.89	60.33
	(5.05)	(4.92)	(4.92)	(4.91)
Inverted	58.00	55.94	53.00	51.61
	(5.50)	(5.59)	(5.80)	(5.75)
Young novices				
Upright	74.86	69.75	61.09	64.60
	(4.08)	(3.97)	(3.97)	(3.97)
Inverted	62.97	61.01	56.88	57.5
	(4.44)	(4.51)	(4.68)	(4.64)

Figure Captions

Figure 1. Example dog stimuli shown here in three of the pairs used for the test phase of the memory task (Experiment 1).

Figure 2. Examples of faces from each of the databases (paired as described in Figure 1).

Figure 3. Experiment 1: Accuracy for the recognition memory test for (a) young adult novice subjects, (b) novices age-sex-education matched to the experts, and (c) expert subjects. Error bars are as appropriate for making the within-subjects comparison between upright and inverted orientations (i.e., ± 1 SEM of the *upright – inverted* difference scores). *** $p < .001$, * $p < .05$, ns = $p > .05$. Note that regarding evidence of basic behavioural expertise (i.e., Are experts better than novices with upright dogs?), only the comparison between experts and age-matched controls is valid (Table 1); see General Discussion for comments on the comparison of experts and young novices.

Figure 4. Experiment 1: Scatter plot of the size of the inversion effect for dogs versus the number of years of experience. Each filled symbol represents one expert. The unfilled diamonds on the left are the averages for young novices (Y) and age-matched novices (AM) ±1 SD.

Figure 5. Experiment 2: Examples of composite face pairs. The target half is the forehead (so the chins are always different-identity). Readers should be able to observe the composite effect for faces on same-identity trials in the left panels. Here, the two top halves are the same identity (although differing in size and brightness), but this is more difficult to see in the aligned condition (top-left) than in the unaligned condition (bottom left).

Figure 6. Experiment 2: Examples of composite dog pairs. The target half is the top (head/tail half).

Figure 7. Experiment 2: Accuracy to compare target half of a pair of composites (simultaneous presentation same-different task) for same trials, collapsed across top-half-to-compare and bottom-half-to-compare blocks. Error bars are as appropriate for the within subjects comparison of aligned and unaligned conditions (i.e., ± 1 SEM of the difference scores). The predicted pattern for a composite effect, indicating holistic/configural processing, is that accuracy should be lower for aligned than unaligned trials. *** $p < .001$, ** $p < .01$, ns = $p > .05$ for aligned vs. unaligned. Note that this experiment cannot be used to assess evidence of basic expertise given age and procedural differences between expert and novice groups (also, the task is to match half dogs, where there is no clear prediction of direction; see General Discussion).

Figure 8. Experiment 2: Scatter plot of the size of the composite effect for dogs versus the number of years experience. Each filled symbol represents one expert. The unfilled diamond on the left represents the average for young novice subjects (±1 SD).

Figure 9. A contrast reversed face and dog as they appeared in Experiment 3.

Figure 10. Experiment 3: Accuracy to judge same or different identity for young novices (A & B) and experts (C & D). Error bars are as appropriate for the within subjects comparison of the two contrast conditions (i.e., ± 1 SEM of the difference scores). *** $p < .001$, ** $p < .01$, ns = $p > .05$ for both original vs. both reversed. Direct comparison of experts to novices cannot be used to assess evidence of behavioural expertise, given age and procedural differences. Comparison of dogs relative to faces within each group is valid (i.e., showing that for upright normal contrast stimuli, experts are as good with dogs as faces while novices are poorer with dogs than faces; see Table 2 and General Discussion).

Figure 11. Experiment 3: Scatter plot of the size of the contrast reversal effect for dogs versus the number of years experience. Formatting as in Figure 8.

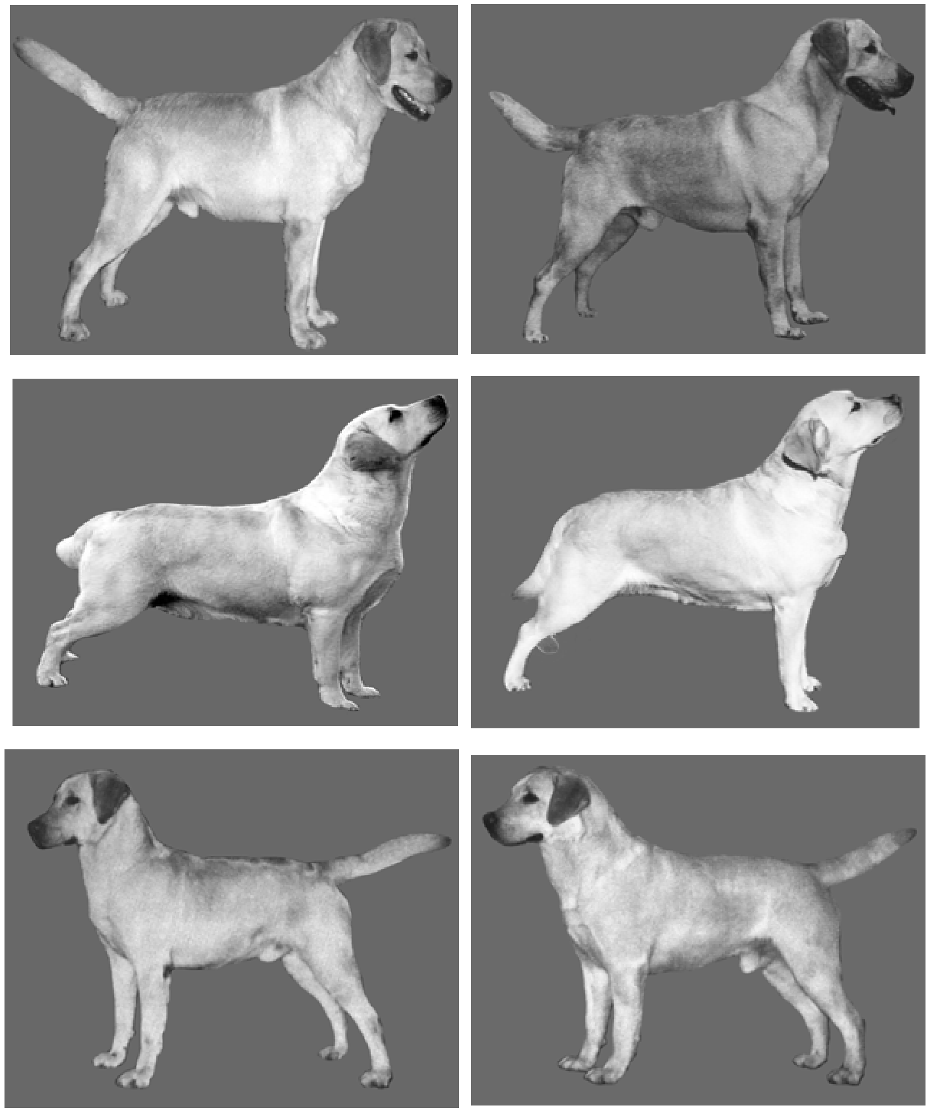

Figure 1. Example dog stimuli shown here in three of the pairs used for the test phase of the memory task (Experiment 1).

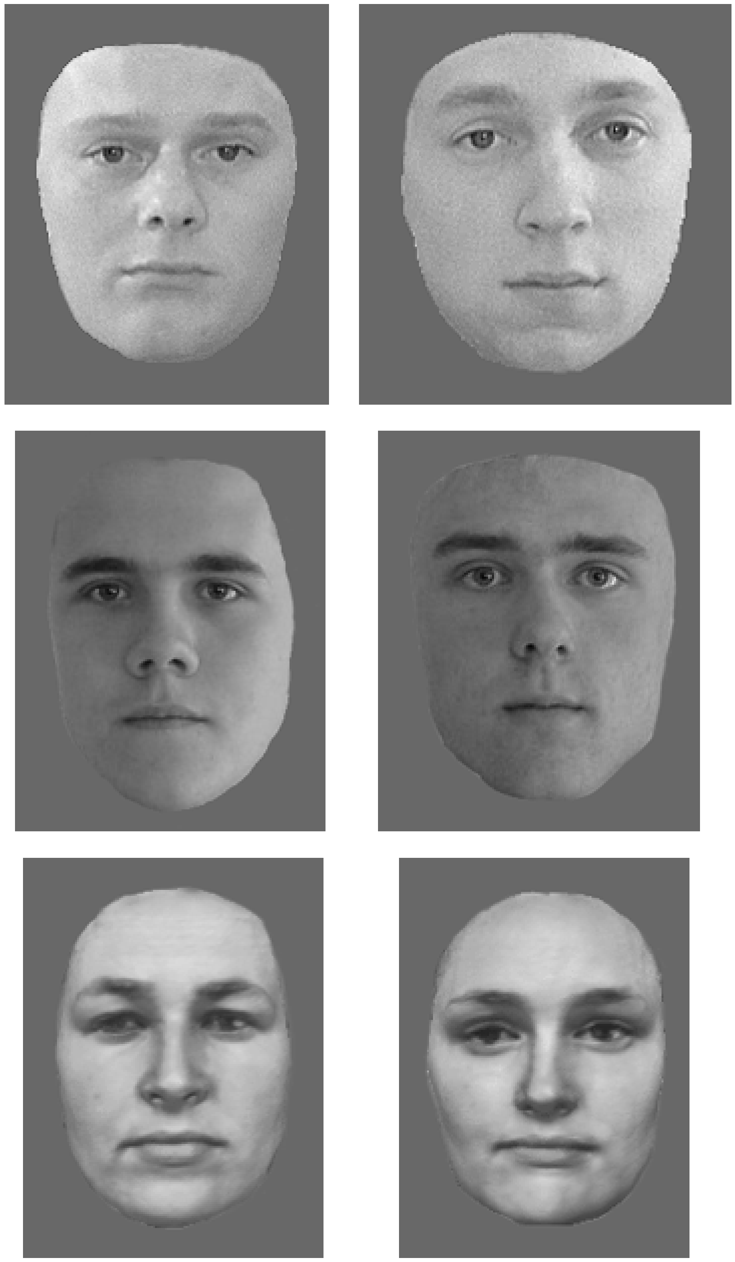

Figure 2. Examples of faces from each of the databases (paired as described in Figure 1).

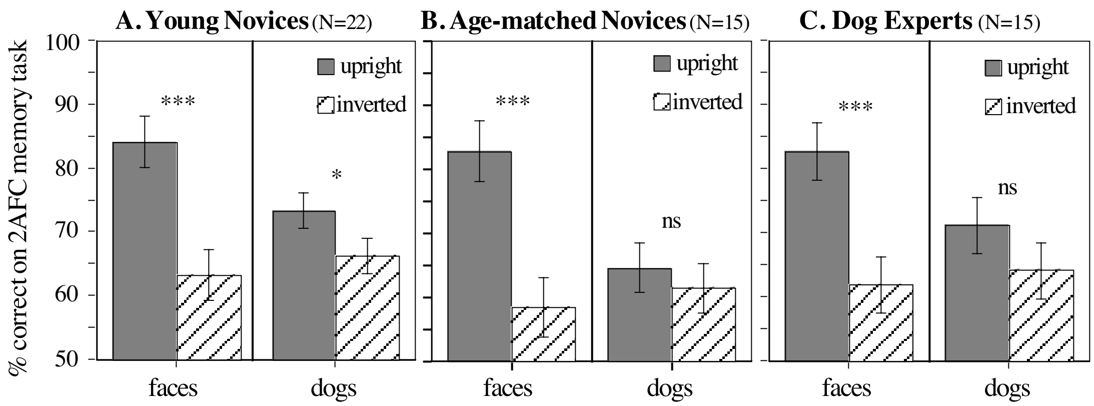

Figure 3. Experiment 1: Accuracy for the recognition memory test for (a) young adult novice subjects, (b) novices age-sex-education matched to the experts, and (c) expert subjects. Error bars are as appropriate for making the within-subjects comparison between upright and inverted orientations (i.e., ± 1 SEM of the *upright – inverted* difference scores). *** $p < .001$, * $p < .05$, ns = $p > .05$. Note that regarding evidence of basic behavioural expertise (i.e., Are experts better than novices with upright dogs?), only the comparison between experts and age-matched controls is valid (Table 1); see General Discussion for comments on the comparison of experts and young novices.

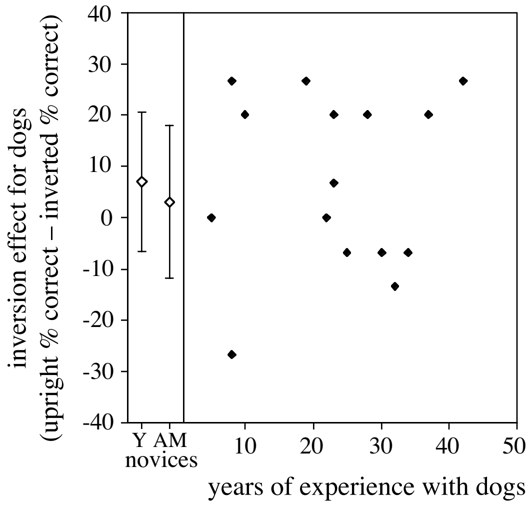

Figure 4. Experiment 1: Scatter plot of the size of the inversion effect for dogs versus the number of years of experience. Each filled symbol represents one expert. The unfilled diamonds on the left are the averages for young novices (Y) and age-matched novices (AM) ±1 SD.

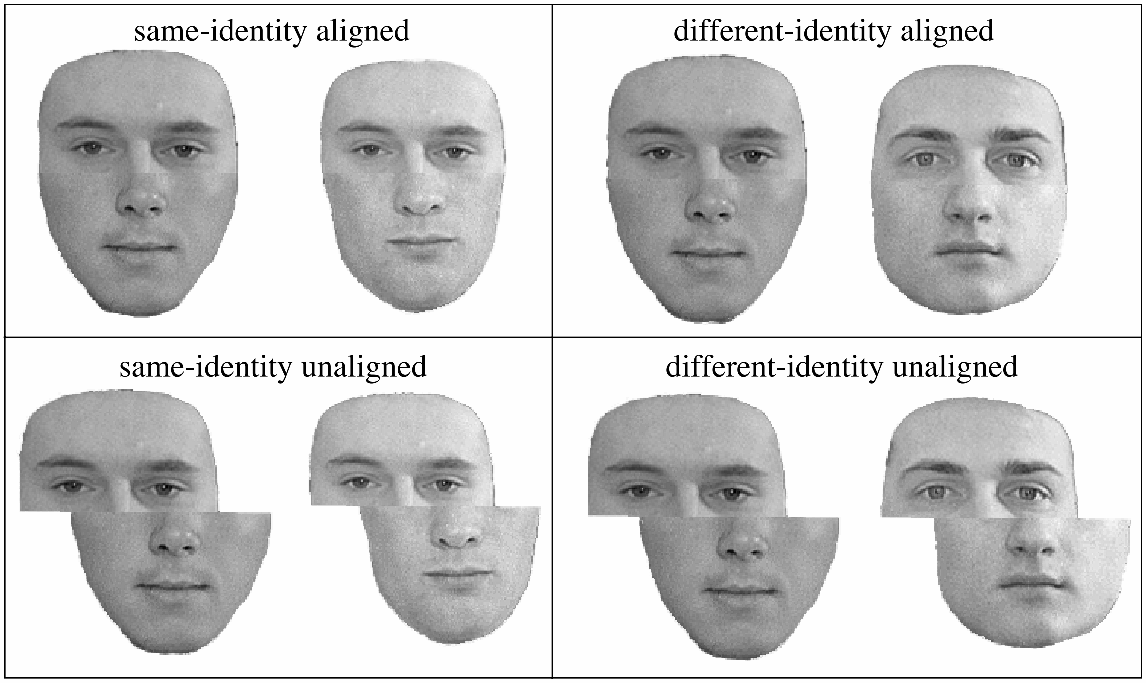

Figure 5. Experiment 2: Examples of composite face pairs. The target half is the forehead (so the chins are always different-identity). Readers should be able to observe the composite effect for faces on same-identity trials in the left panels. Here, the two top halves are the same identity (although differing in size and brightness), but this is more difficult to see in the aligned condition (top-left) than in the unaligned condition (bottom left).

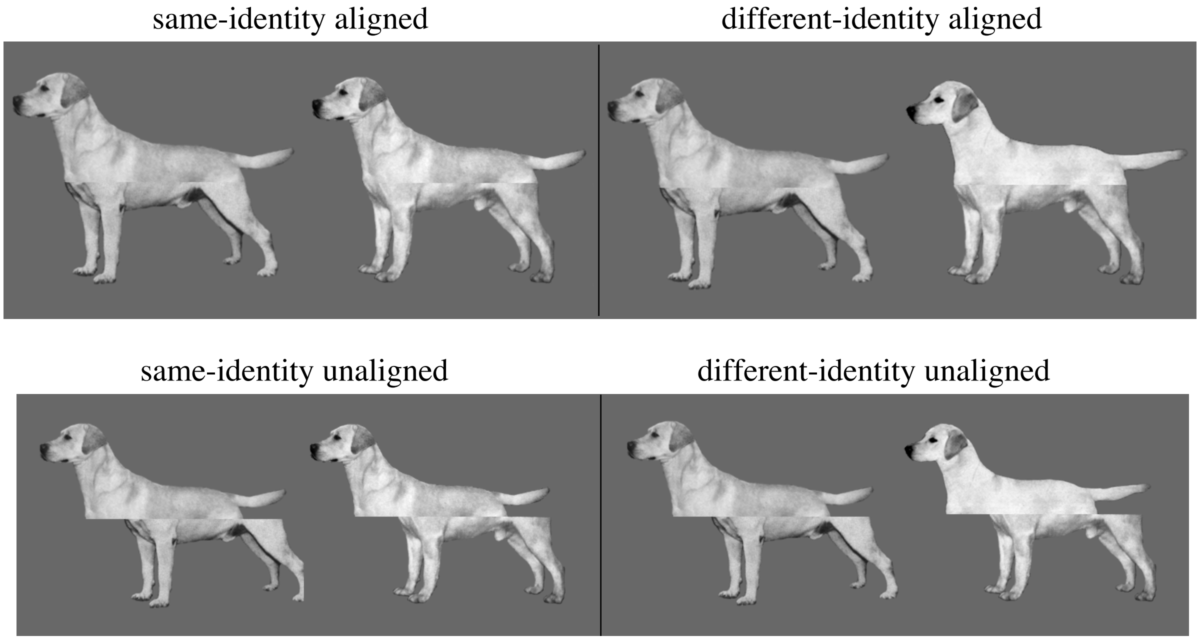

Figure 6. Experiment 2: Examples of composite dog pairs. The target half is the top (head/tail half).

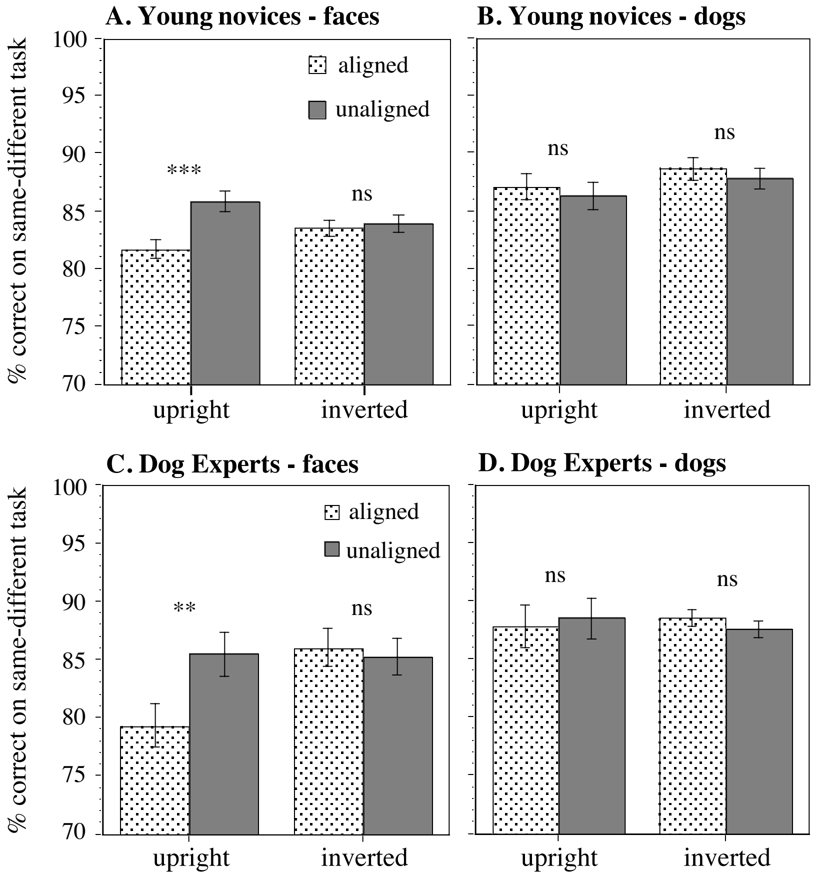

Figure 7. Experiment 2: Accuracy to compare target half of a pair of composites (simultaneous presentation same-different task) for same trials, collapsed across top-half-to-compare and bottom-half-to-compare blocks. Error bars are as appropriate for the within subjects comparison of aligned and unaligned conditions (i.e., ± 1 SEM of the difference scores). The predicted pattern for a composite effect, indicating holistic/configural processing, is that accuracy should be lower for aligned than unaligned trials. *** $p < .001$, ** $p < .01$, ns = $p > .05$ for aligned vs. unaligned. Note that this experiment cannot be used to assess evidence of basic expertise given age and procedural differences between expert and novice groups (also, the task is to match half dogs, where there is no clear prediction of direction; see General Discussion).

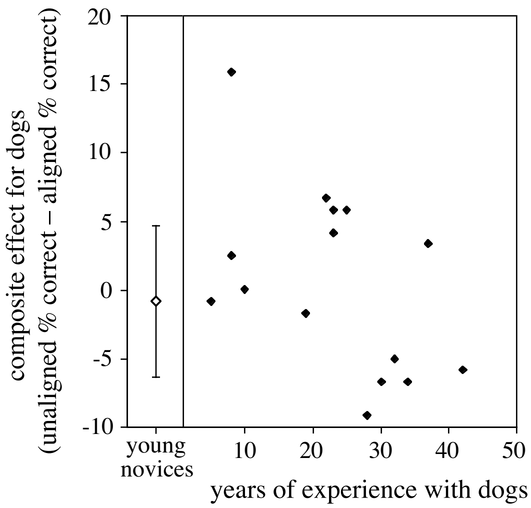

Figure 8. Experiment 2: Scatter plot of the size of the composite effect for dogs versus the number of years experience. Each filled symbol represents one expert. The unfilled diamond on the left represents the average for young novice subjects (±1 SD).

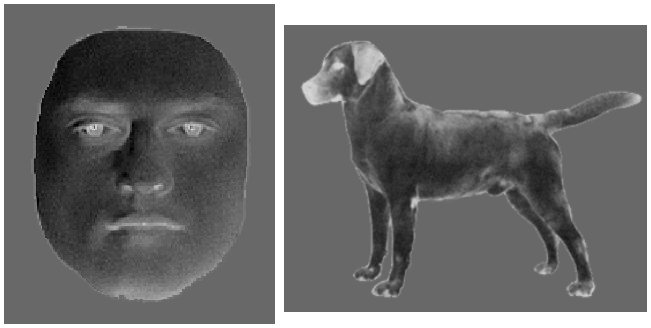

Figure 9. A contrast reversed face and dog as they appeared in Experiment 3.

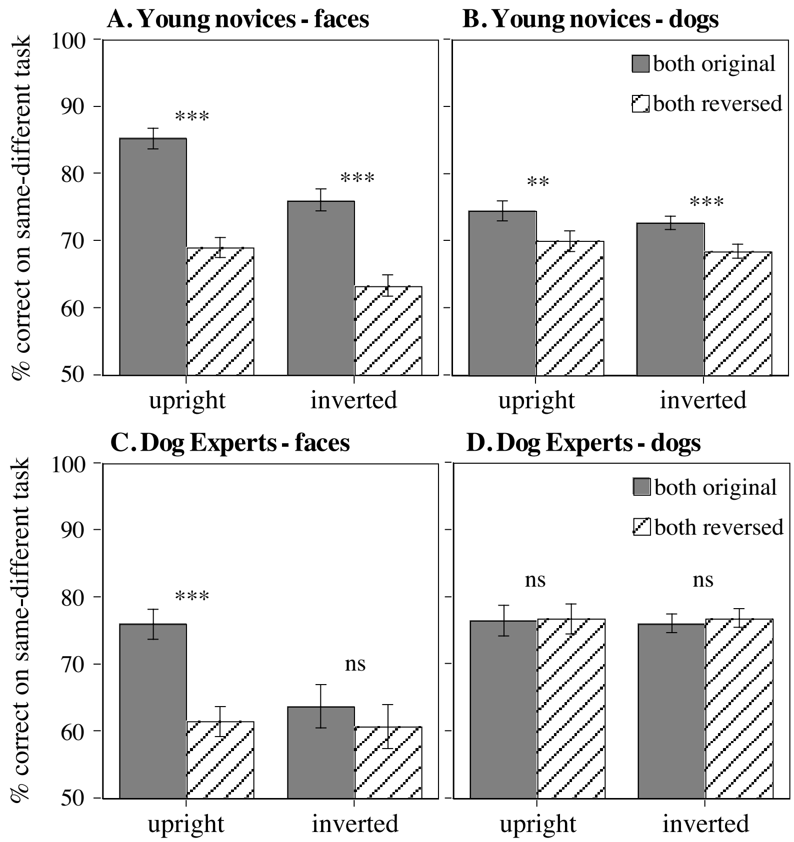

Figure 10. Experiment 3: Accuracy to judge same or different identity for young novices (A & B) and experts (C & D). Error bars are as appropriate for the within subjects comparison of the two contrast conditions (i.e., ± 1 SEM of the difference scores). *** $\underline{p} < .001$, ** $\underline{p} < .01$, ns = $\underline{p} > .05$ for both original vs. both reversed. Direct comparison of experts to novices cannot be used to assess evidence of behavioural expertise, given age and procedural differences. Comparison of dogs relative to faces within each group is valid (i.e., showing that for upright normal contrast stimuli, experts are as good with dogs as faces while novices are poorer with dogs than faces; see Table 2 and General Discussion).

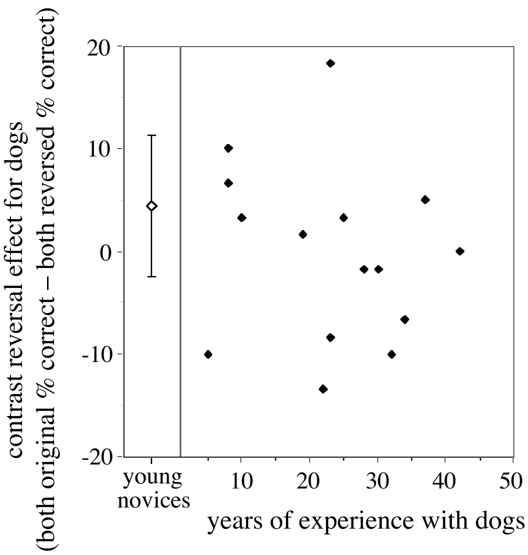

Figure 11. Experiment 3: Scatter plot of the size of the contrast reversal effect for dogs versus the number of years experience. Formatting as in Figure 8.