10

Face and Object Recognition
How Do They Differ?

ELINOR MCKONE

INTRODUCTION

This chapter reviews an extensive set of findings arguing that visual recognition of faces and objects differs critically in style of computational processing. It provides a tutorial-style review of what face-recognition researchers mean by "holistic/configural" processing and describes the multiple paradigms showing that holistic/configural processing is limited to the structural form of upright faces. It then brings together several streams of literature to present a theoretical case that holistic/configural processing and part-based processing differ in patterns of sensitivity to prior experience.

Findings show holistic/configural processing is not learnable for objects, even with expertise; is insensitive to amount of experience with different viewpoints of faces; and (contrary to early ideas) does not require many years of exposure to develop in childhood. Holistic/configural processing is sensitive to experience only in that, for *faces as an entire class*, it has a critical period in infancy and that, for particular *subtypes* (e.g., races) of upright faces, it can weaken or strengthen throughout life. In contrast, part-based processing, as occurs for both faces and objects, is strongly sensitive to experience. It improves rapidly, even with experimental practice, and has no critical period in infancy for exposure to particular structural forms.

RATIONALE

My research interests lie primarily in face recognition, particularly the perceptual processes involved in face identification. In attempting to understand these processes, a fundamental question is: Are faces different from objects? The answer to this question is crucial. Without it, we do not know, for example, whether computational and neural models of face recognition must be general enough to perform recognition of all other objects also or can be limited to the (presumably easier) task of coding the structural form of faces alone. As another example, we do not know whether we should be looking for different evolutionary mechanisms to drive the visual processes of face and object recognition—perhaps deriving from their different levels of social importance or their different roles in social communication. Should we instead be searching for a common evolutionary mechanism that might drive recognition of complex visual stimuli in general?

This chapter presents a strong thesis in answer to this question. I argue that visual recognition of faces and objects differs in at least two ways: (a) the style of computational processing used to identify them, and (b) the patterns of sensitivity of that style of processing to prior experience. Specifically, I argue that *holistic/configural* processing occurs for faces but not objects, and that this "special" style of processing is insensitive to many types of adult experience (e.g., practice with objects or with unusual views of faces), although experience is required during a critical period in infancy. In contrast, part-based processing is the means for identifying objects, and this is highly sensitive to experience and can be easily learned for new classes of object even as an adult.

Note that these differences are located at the stage of "high-level perception" or "visual recognition" (occurring in inferior and lateral regions of the temporal lobe) and should not be taken to imply a claim that face and object processing will be different in all possible ways. Many other stages of the processing stream will occur in common for both stimulus types, such as early visual processing, post-recognition cognitive decision mechanisms, possible involvement of working memory in the task, and so on. It is also very likely that face and object systems share some operational principles in common; for example, all visual recognition systems show frequency effects in which reaction times are faster and functional magnetic resonance imaging (fMRI) BOLD response is lower to high-familiarity as compared to low-familiarity items.

This chapter also has several other aims. I present a somewhat tutorial-style review of the paradigms dissociating faces from objects that is designed to be useful to researchers new to the field of behavioral studies of holistic/configural processing. For readers expert in this area, certain material in the review section—particularly a discussion of the dissociation between coding of spacing and local feature information—also leads to a more detailed theoretical discussion of the nature of holistic/configural processing for faces than is possible in the standard empirical research article. This includes a critique of the different terminologies that face-recognition researchers have used (e.g., holistic vs. configural), an evaluation of Maurer, Le Grand, and Mondloch's (2002) proposal that there are subcomponents to configural processing, and an explanation of my rather different theoretical position.

Finally, I discuss the possible *origins* of special processing for faces. I expand on theoretical ideas we presented in McKone, Kanwisher, and Duchaine (2007). I consider the ability of an *experience-expectant innate template* theory and an *infant experience plus other factor* theory to explain not only the holistic/configural processing findings, but also other key observations about face recognition such as the heritability of developmental prosopagnosia. Readers should note this chapter was finalized in August, 2007, and that only research available up until that date is reviewed.

A DIFFERENT COMPUTATIONAL STYLE FOR RECOGNIZING FACES AND OBJECTS: HOLISTIC AND PART-BASED PROCESSING

In the context of the face-recognition literature, holistic/configural processing is thought to involve especially strong integration of information across the whole face region (excluding hair), which occurs at a perceptual level for stimuli that require individuation based on second-order deviations from a shared first-order configuration. First-order structure is defined as standard parts—eyes, nose, etc.—in the standard configuration of eyes above nose above mouth. By "second-order information" I mean individual exemplar deviations from this structure—for example, in exact distances between features or in exact feature shape. In the extreme view (e.g., Tanaka & Farah, 1993), holistic/configural processing comprises no decomposition into smaller parts at all, although another common idea is that it is based on some type of particularly strong reintegration following initial part decomposition.

Holistic/configural processing is usually contrasted with part-based, local, componential, or featural analysis, where parts of an object (or a face) are treated relatively independently. Empirical evidence indicates that inverted faces, scrambled faces, and objects are processed in a part-based fashion, while only upright faces are processed holistically/configurally. Evidence also indicates that upright faces receive both configural processing *and* part-based processing, rather than *only* configural processing. Thus, on many behavioral tasks, performance with upright faces will be based on a combination of the two processing types.

I now review results from relevant empirical paradigms. Except where otherwise mentioned, all results to be described come from studies using realistic face stimuli (usually grayscale photographs). Another important fact to note is that all the results to be described come from "ordinary" people—that is, typical people who are very good at recognizing individual faces but have no special expertise with the other object class or classes included in the experiment. In the literature, such people are known as object *novices*. They may have general familiarity with the object class, in its canonical orientation, but that is all. For example, most of us are familiar with upright dogs but are poor at telling apart individual dogs of the same breed. A later section deals with the topic of object experts.

Classic Holistic/Configural Paradigms

Several standard findings have been associated with holistic/configural processing. The first is the *disproportionate inversion effect* (Yin, 1969). In both recognition memory and perception (simultaneous or sequential matching tasks), all objects with a canonical upright show an inversion effect: Accuracy is higher and reaction time faster when the stimulus presentations are all upright than when they are all inverted (upside-down). However, the inversion effect for most objects is small, typically ranging from 0 to 8% in any one recognition memory experiment for dogs, cars, clothing, and so on. In contrast, the inversion effect for faces is very large, typically around 20–25% in a recognition memory experiment. This result was first reported by Yin (1969) and has subsequently been confirmed by many other studies (e.g., see first two data columns of Table 10.1; also see de Gelder, Bachoud-Levi, & Degos, 1998; Reed, Stone, Bozova, & Tanaka, 2003; Scapinello & Yarmey, 1970).

Yin's (1969) original interpretation of his results was that extracting the correct relationships between the face parts was particularly important to face recognition and that extracting this information from inverted faces was difficult. Note, however, that the disproportionate inversion effect itself provided no direct evidence for holistic/configural processing; logically, it could also have been that the style of processing for faces and objects was the same (e.g., part based in both cases) and it was merely that inversion effects on this common style were largest for faces. Similar arguments were made by Valentine (1988, 1991). Today, the disproportionate inversion effect is referred to as "indirect" evidence for holistic processing for upright faces (e.g., Michel, Rossion, Han, Chung, & Caldara, 2006). Maurer et al. (2002) explicitly noted that the mere presence of an inversion effect is not diagnostic of holistic/configural processing.

Three paradigms were subsequently designed that assess processing style more directly. These all confirm differences between upright faces and inverted faces, and between faces and objects. In Tanaka and Farah's (1993) *part–whole paradigm*, after learning whole faces ("This is Jim."), subjects are given a 2AFC recognition memory test (see Figure 10.1). In the part-alone condition, the subject sees, for example, Jim's nose and a distractor person's nose and is asked to choose Jim's nose. In the part-in-whole condition, exactly the same parts are presented, but now both are shown in the context of Jim's face (e.g., Jim's nose in Jim's face vs. Bill's nose in Jim's face). The task is either "Which is Jim?" (Tanaka & Farah, 1993) or "Which is Jim's nose?" (Tanaka & Sengco, 1997).

Results from multiple studies show that, despite the fact that the physical difference between the pairs is exactly the same in both conditions, identification is substantially better in the part-in-whole condition than in the part-alone condition (see Table 10.2; also see Pellicano & Rhodes, 2003; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998). This corresponds to the illusion, visible in Figure 10.1, that changing one feature of the face alters the appearance of the rest of the face.

Importantly, the part–whole effect disappears for inverted faces; that is, the part-in-whole condition has no advantage over the part-alone condition (Table 10.2; Pellicano & Rhodes, 2003). Readers can appreciate this for themselves by turning Figure 10.1 upside down to see the corresponding lack of illusion. The part–whole effect has also been tested for objects, using both the Tanaka and Farah (1993) procedure and a very similar earlier procedure (*complete probe advantage*; Davidoff & Donnelly, 1990). For line drawings of houses, which one might argue are particularly easily separable into parts, there is no part–whole effect (Tanaka & Farah, 1993). For other types of objects, the part–whole effect is greater than zero but much smaller than for faces (see Table 10.2; also see Donnelly & Davidoff,

TABLE 10.1 Inversion Decrement (Upright–Inverted) for Faces and Objects

	Task	Faces	Objects (novices)	Objects (experts)	Sig. of Expertise Increase
Dogs (Diamond & Carey, 1986, Experiment 3)	Long-term memory	20%[*]	5%[ns]	22%[*]	*
Dogs (Robbins & McKone, 2007, Experiment 1)	Long-term memory	23%[*]	3%[ns]	7%[ns]	ns
Handwriting (Bruyer & Crispeels, 1992)	Long-term memory	20%[*]	5%[ns]	9%[*]	ns
Dogs (Robbins & McKone, 2007, Experiment 3)	Sequential matching	11%[*]	1%[ns]	2%[ns]	ns
Greebles (Rossion et al., 2002)	Sequential matching	75 ms[−]	25 ms[−]	46 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%[−]	8%[−]	ns
Fingerprints (Busey & Vanderkolk, 2005)	Face/print classification	– 16%[*]	d' = 0.44[−] 6%[ns]	d' = 0.87[−] 8%[ns]	* ns

Notes: Because there are too many studies testing faces versus objects in novices to list all of them, only studies that also tested experts on the objects are shown; all such studies are included. Studies reported various measures, including percent correct (%), d', and reaction time (milliseconds). 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. Greebles are an artificial object class. Means for Busey & Vanderkolk (2005) provided by Thomas Busey (pers comm, July 21, 2006); percent correct means for Xu et al. (2005) provided by Yaoda Xu (pers comm, August 25, 2004). Adapted and expanded from Robbins R., & McKone, E., 2007, *Cognition, 103*, 34–79.

[*] $p < .05$.

[ns] $p > .05$.

[−] Not tested or not reported.

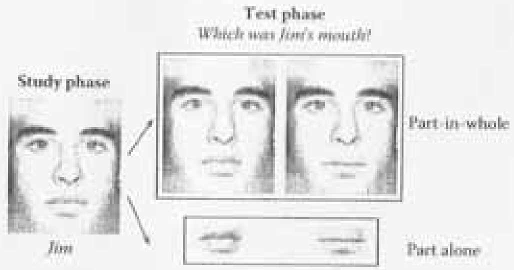

Figure 10.1 The part–whole procedure of Tanaka & Farah (1993, *Quarterly Journal of Experimental Psychology A, 46*, 225–245) and the corresponding perceptual illusion. When the face is upright, changing one feature of the face gives the illusion of changes to nonaltered regions: Replacing Jim's original mouth with one that has narrower lips increases the lip–nose distance and produces an impression of a shorter nose and more squished-up eye–nose region. The illusion disappears if the page is rotated to see the stimuli inverted.

1999). These results are consistent with the idea that what makes upright faces special is holistic/configural processing.

A number of criticisms of the part–whole method have claimed that it taps something other than perceptual holistic integration. Gauthier and Tarr (2002) suggested that, because it occurs for objects, it merely assesses an advantage of context that is generic to all stimuli. I agree that there may be some generic context component to the part–whole advantage, presumably reflecting the same sort of process that also occurs in the classic word superiority effect (Reicher, 1969; Wheeler, 1970). However, this cannot be all that there is because it fails to explain why the effect is so much larger for faces than for other objects.

Another idea is that the effect reflects merely the well established phenomenon of transfer-appropriate processing (TAP; or encoding specificity) in long-term memory; that is, memory retrieval is better in the whole condition at test because this matches the conditions at encoding, which also showed a whole face. Leder and Carbon (2005) found that the usual whole-over-part advantage was reversed when the study phase involved learning parts rather than whole faces. Again, however, this merely indicates that the part–whole effect—in the long-term memory version of the paradigm at least—does contain some component due to TAP. It does not explain why the effect is so much larger for faces than for other objects. The study–test match in the whole condition is the same for all stimulus classes.

The next paradigm is a variant on the part–whole effect, which has gone under various names but which I refer to here as the *part in spacing altered whole* effect. This was introduced by Tanaka and Sengco (1997). For faces, memory for a face feature (Jim's nose) is worse in the context of a spacing-altered version of the original face (Jim's nose in Jim's face with the eyes shifted further apart) than in the unaltered version (Jim's nose in Jim's face). This result is important in that it demonstrates that the usual part–whole advantage for faces does not come merely from the presence of extra context provided by having more features in a part-in-whole

TABLE 10.2 Results of Previous Studies, Using the Tanaka & Farah (1993) Part–Whole Paradigm, Showing Size of the Whole–Part Difference, Averaged Over All Parts Tested

	Faces	Inverted Faces	Objects (novices)	Objects (experts)	Sig. of Expertise Increase
Houses (Tanaka & Sengco, 1997)	11%*	−1%**	−2%**	–	–
Houses (Tanaka & Sengco, 1997)	15%*	0%**	1%**	–	–
Chairs (Davidoff & Donnelly, 1990)	11%*	–	4%**	–	–
No objects (Pellicano, Rhodes, & Peters, 2006)	13%*	−2%**	–	–	–
Dog faces (Tanaka et al., 1996)	20%*	–	2%**	8%**	ns
Cars (Tanaka et al., 1996)	18%*	–	8%*	6%*	Reverse
Biological cells (Tanaka et al., 1996)	26%*	–	16%*	10%*	Reverse
Greebles (Gauthier & Tarr, 1997)	–	–	5%**	11%*	ns
Greebles (Gauthier et al., 1998)	–	–	7%*	0%*	Reverse
Greebles (Gauthier & Tarr, 2002)	–	–	$d' = 0.75$*	$d' = 0.68$*	Reverse

Notes: All stimuli were upright unless otherwise stated. All studies including objects of expertise are included, as are sample extras that tested objects only in novices. Adapted and expanded from Robbins R., & McKone, E., 2007, *Cognition, 103*, 34–79. Data from Tanaka et al. (1996) are as cited in Tanaka, J. W., & Gauthier, I., 1997, 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.
* $p < .05$.
** $p > .05$.
– Not tested or not reported.
Reverse = Trend in opposite-to-predicted direction for expertise effect.

TABLE 10.3 Results of Previous Studies, Using Tanaka & Sengco's 1997 Paradigm, Showing *Part-in-Whole* Minus *Part-in-Spacing-Altered-Whole*, Averaged Over All Parts Tested

	Faces	Objects (novices)	Objects (experts)	Sig. of expertise increase
Houses (Tanaka & Sengco, 1997)	7%*	0%**	–	–
Greebles (Gauthier & Tarr, 1997)	–	–4%**	0%**	ns
Greebles (Gauthier et al., 1998)	–	1%**	0%**	Reverse
Greebles (Gauthier & Tarr, 2002)	–	$d' = 0.60$*	$d' = 0.64$*	Reverse

Notes: All stimuli were upright unless otherwise stated. All studies that tested objects of expertise are included, as is the original study that tested objects only in novices.
* $p < .05$.
** $p > .05$.
– Not tested or not reported.
Reverse = Trend in opposite-to-predicted direction for expertise effect.

condition than in the part-alone condition. Instead, it depends on reinstating the specific arrangement of studied features. Empirical findings indicate that the advantage of whole over configurally transformed whole is strong for upright faces and absent for inverted faces (Pellicano, Rhodes, & Peters, 2006; Tanaka & Sengco, 1997); for objects it is either absent (Gauthier, Williams, Tarr, & Tanaka, 1998; Tanaka & Sengco, 1997) or weak (Gauthier & Tarr, 1997, 2002; see Table 10.3).

The next task is the *composite paradigm* of Young, Hellawell, and Hay (1987), illustrated in Figure 10.2. In the famous-face version, composites are formed of the top half of one individual and the bottom half of a different individual, and the task is to name one half (e.g., the top half). The halves are presented aligned or unaligned. In the former case, an illusion of a new person is created. Correspondingly, it is substantially harder (i.e., slower and/or less accurate) to name the target half in the aligned version than in the unaligned version.

However, as usual, this effect occurs only for upright faces rather than for inverted faces (Young et al., 1987). The same result is found in versions of the paradigm using initially unfamiliar faces that subjects are trained to name (Carey & Diamond, 1994; McKone, 2008; Robbins & McKone, 2003) and for completely novel faces in sequential or simultaneous same–different matching (de Heering, Houthuys, & Rossion, 2007; Le Grand, Mondloch, Maurer, & Brent, 2004; Michel, Rossion, et al., 2006; Robbins & McKone, 2007).

In the last few years, the composite paradigm has become increasingly popular as a good method to measure holistic processing. Logically, it has an advantage over the part–whole paradigm in that the same amount of information is present

Figure 10.2 The composite procedure of Young et al. (1987, *Perception, 16*, 747–759) using famous faces and the corresponding perceptual illusion. When the face is upright, aligning the two halves leads to a percept of a new individual and to corresponding difficulty (increased reaction time and/or decreased accuracy) in naming a target half; in contrast, the identity of each half is easy to see in the unaligned version. When the stimuli are inverted, the illusion disappears. For answers, see the "Acknowledgments" section.

on the screen in the two conditions compared (aligned and unaligned). Also, the presence of a competing response suggested by the nontarget half (e.g., the name if the face is familiar or the different-identity status if it is novel) is the same in both conditions.

Empirically, the composite effect does not reflect merely a general failure to tune out the nontarget half of the face: If this were the case, attentional competition from a (closer) aligned half would be stronger than from a (further away) unaligned half and thus the effect should be obtained for inverted faces, which it is not. Finally, the composite effect appears to show the clearest dissociation between faces and objects. The inversion and part–whole paradigms both produce a partial dissociation in which the target effect is much larger for faces than objects but is still present at some level for objects. The composite effect, however, appears to be genuinely absent for objects in novices (see Table 10.4), suggesting that it provides a purer measure of holistic/configural processing than other paradigms.

The Special Case of Spacing-Versus-Feature Changes

Another very standard approach to investigating configural/holistic processing, usually used for faces rather than objects, has been to make alterations to the face either in the distances between facial features (e.g., moving the eyes further apart) or in the appearance of individual features (e.g., changing the eyebrows). It has been common to associate sensitivity to spacing changes with configural/holistic processing and sensitivity to feature changes with part-based processing (e.g., Maurer et al., 2002). However, I think the situation is more complex than this.

The evidence for the standard idea comes from many studies that have found dissociations in the effect of inversion. Inversion influences perception of spacing or relational changes particularly strongly. This was originally discovered in

TABLE 10.4 Results of the Young & Colleagues' 1987 Composite Paradigm, Showing the Aligned–Unaligned (for Reaction Times) or Unaligned–Aligned (for Accuracy) Difference

	Task	Faces	Inverted faces	Objects (novices)	Objects (experts)	Sig. of expertise increase
No objects (Young et al., 1987)	Speeded naming	212 ms*	9 ms**	–	–	–
No objects (Robbins & McKone, 2003)	Naming twins	8.8%*	–1.2%**	–	–	–
Greebles, same-family halves (Gauthier et al., 1998)	Speeded naming	–	–	–	115 ms** 0%	–
Greebles, different-family halves (Gauthier et al., 1998)	Speeded naming	–	–	–	–37 ms[b,**] –3%[b,**]	– –
Greebles, same-family halves (Gauthier & Tarr, 2002)	Speeded naming	–	–	–42 ms*	12 ms*	–[a]
Dogs (Robbins & McKone, 2007)	Sequential matching	6.1%*	–3.5%**	–0.8%** 0.8%[**,b]	0.7%**	ns Reverse

Notes: In both the aligned–unaligned and unaligned–aligned 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 unless otherwise stated. All studies including objects-of-expertise are included, as are some sample studies that reported data for inverted faces. Adapted and expanded from Robbins R., & McKone, E., 2007, *Cognition, 103*, 34–79.

* p < .05.

** p > .05.

– Not tested or not reported.

Reverse = Trend in opposite-to-predicted direction for expertise effect.

[a] Across five sessions (we show only: session 1 = novices; session 5 = experts), there was a close-to-significant interaction between session and aligned versus 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, and then returned to close to zero. The 12-ms composite effect in experts was in the context of 35-ms scanning electron microscopes for the aligned and unaligned conditions. Also note that Gauthier et al. (1998, *Vision Research*, 38, 2416) mention two additional earlier failed attempts to find a composite effect using the greeble experts from Gauthier & Tarr (1997, *Vision Research*, 37, 1673–1682).

[b] Results for two independent groups of subjects.

the famous Margaret Thatcher illusion (Thompson, 1980), in which flipping the eyes and mouth of a smiling face makes the face appear grotesque when the head is upright, but not when it is inverted. Rhodes, Brake, and Atkinson (1993; also see Bartlett & Searcy, 1993) demonstrated that, in faces made bizarre by spacing changes (e.g., moving the mouth down), inverting the face markedly reduced the perception of bizarreness. Inversion had only a weak effect, however, on perception of faces made bizarre by local feature changes (e.g., blackening the teeth).

Leder and Bruce (1998; also see Gilchrist & McKone, 2003) made less severe changes that merely made faces look more or less distinctive. Spacing changes that increased perceived distinctiveness and improved recognition memory for upright faces had weak or no effects for inverted faces, but featural changes had equally strong effects in both orientations. Using sequential same–different tasks, simple detection of spacing changes has also been found to be more severely affected by inversion than detection of featural changes (Freire, Lee, & Symons, 2000; Le Grand, Mondloch, Maurer, & Brent, 2001; Mondloch, Dobson, Parsons, & Maurer, 2004; Mondloch, Le Grand, & Maurer, 2002).

How true is the claim that feature changes show only small inversion effects? Riesenhuber, Jarudi, Gilad, and Sinha (2004) reported that inversion effects on detecting feature changes were larger when feature and spacing changes were intermixed than when they were blocked; more generally, they argued that findings of no inversion effects for featural changes were due to subjects adopting unusual strategies in blocked procedures. The latter claim, however, ignored the fact that Leder and Bruce (1998) and Gilchrist and McKone (2003) both found no inversion effect on feature change tasks using mixed presentation.

In another critique, Yovel and Kanwisher (2004) noted that Le Grand et al. (2001) and Mondloch et al. (2002, 2004) had all used the same stimulus set and that, unfortunately, this set was flawed in that the featural changes were easier to perceive upright than the spacing changes were. When Yovel and Kanwisher matched the two types for detectability in the upright orientation, the effects of inversion were equally severe for each. A criticism of the Yovel and Kanwisher (2004) procedure is that, in order to match quite large feature changes, the spacing-changed faces became quite distinctive in appearance (bordering on becoming abnormal), while the feature-changed faces were more typical.

Valentine (1991) has previously shown that distinctiveness can influence the size of inversion effects. In addition, of the previous studies, not all had feature changes that were perceptually larger than the spacing changes. Gilchrist and McKone (2003) had the reverse pattern for upright faces and yet still found that inversion influenced perception of spacing changes but not of featural changes.

Although I am not convinced by the particular explanations offered in Riesenhuber et al. (2004) or Yovel and Kanwisher (2004), the importance of their results cannot be ignored. Both studies show that large inversion effects for feature changes can occur. The same finding has also been reported by Rhodes, Hayward, and Winkler (2006); McKone and Boyer (2006); Leder and Carbon (2006); and Yovel and Duchaine (2006). This substantially undermines the supposed dissociation between spacing and features.

Taking recent findings into account, the standard description of the data needs to be changed, as follows:

> Perception of *spacing* changes is always strongly influenced by inversion.
> The effect of inversion on *feature* changes is much more variable; some studies find no inversion effect, some find a moderate inversion effect, and at least two find an inversion effect as big as that for spacing changes.

How can the apparently conflicting results be understood? Recent evidence suggests that results for the feature-altered faces depend on the types of feature changes. Some changes are purely "local," such as altering eye color, blackening teeth, or adding a long, thick hair to an eyebrow. Others alter the form or structure of surrounding regions of the face; for example, if a large eye is replaced with a smaller eye, then even if the new eye remains centered in the same position as the old eye, the distance from the inside edge of the eye to the bridge of the nose changes, as does the apparent shape of the eye socket.

Two studies demonstrate that shape-altering changes produce larger inversion effects than purely local changes. Yovel and Duchaine (2006) found that feature changes produced large inversion effects when the changes were in shape only (not color) and small inversion effects when changes were in shape plus color. Leder and Carbon (2006) reported larger inversion effects for feature shape replacements (eyes, nose, or mouth) than for color-only changes (with spacing changes producing a larger effect).

Overall, I agree with the standard idea that sensitivity to detailed spacing information forms a key aspect of configural processing. However, I do not agree with the idea implicit in much of the literature that *only* spacing information is part of configural/holistic processing and that any type of feature change is processed in a part-based fashion. This idea is apparent in a tendency to equate second-order information only with distances between the major features (e.g., nose–mouth distance), as if features can be treated as shapeless blobs.

Other Paradigms Consistent With Configural Processing for Upright Faces Only

The paradigms reviewed so far are widely considered the "core" paradigms. They have been used in multiple studies originating from many independent labs and are widely cited. However, many other tasks have also produced evidence consistent with the idea that configural processing occurs only for upright faces. In some cases, these tasks have also been used to test objects; in other cases, the comparison stimuli have been only inverted faces (or similar—for example, single isolated face features).

In the most abstract approach, Loftus, Oberg, and Dillon (2004) used dimensional theory and state-trace plots to investigate the disproportionate inversion effect for faces as compared to houses and cityscapes. Results showed that when proportion-correct memory for famous faces was plotted against proportion-correct memory for houses as a function of stimulus duration at study, upright and inverted data points fell on different functions. Loftus et al. demonstrated that the data were neatly described by a model in which the three independent variables (stimulus class, orientation, and stimulus duration) affected performance via their effects on two internal variables—presumably corresponding to the processing modes of holistic/configural and featural processing.

Using a *flanker variant of the part–whole paradigm*, Palermo and Rhodes (2002) found that a secondary task of matching upright flanker faces removed the part–whole effect for central upright faces. However, matching inverted flankers did not remove the part–whole effect for central upright faces. This argues that upright faces compete with other upright faces for holistic/configural resources, but that inverted faces do not; they are processed by other, presumably part-based resources.

McKone and Peh (2006) used a *memory conjunction procedure*. Test faces were old (unaltered study faces), new (completely unstudied faces), or a new conjunction of old parts. Conjunctions contained the eyes and eyebrows of one studied face combined with the nose and mouth of a different studied face. In a long-term memory test, subjects were required to say whether they had seen the face before. For both upright and inverted faces, the percentage of "old" responses to conjunctions was higher than the false alarm rate for new faces, indicating some memory of isolated face parts (also consistent with above-chance memory for single face parts in Tanaka & Farah, 1993).

Configural processing was then indicated by comparing truly old faces with conjunction faces. For upright, the percentage of "old" responses was much higher for old than for conjunction stimuli; for inverted, the two did not differ. Thus, subjects remembered which parts had been paired together for upright faces—consistent with holistic processing—but did not for inverted faces—consistent with part-based processing.

Cohen and Cashon (2001) used a related procedure in infants. Infants habituated to two female faces, then saw one of the original faces, a new face, or a conjunction formed from the internal features of one original (eyes, nose, and mouth) with the outer regions (hair, cheeks, and chin) of the other. Infants treated the conjunction face as new when the faces were upright (dishabituation to a new relationship between old parts) but as old when they were inverted (no dishabituation; i.e., the new relationship between parts was ignored).

In other approaches to demonstrating *interaction between parts*, Sergent (1984) showed subjects six line-drawn face stimuli varying on three dimensions (external contour shape, eyes/eyebrows shape, and nose–mouth spacing) of two values each. Regression analyses predicting errors and reaction times on a simultaneous matching task showed that the manipulated features contributed interactively when the faces were upright, but only independently when the faces were inverted. In Sergent's second experiment, a multidimensional scaling analysis of dissimilarity judgments between pairs of the upright faces revealed similar findings.

Yovel, Paller, and Levy (2005) tested subjects on various combinations of left or right hemifaces. The test stimulus on each trial comprised a choice of six faces, all bilaterally symmetric (i.e., the same individual on both halves). The briefly presented and masked study stimulus was a whole face showing the same individual in both halves, a whole face showing different individuals in the two halves, a stimulus showing the left hemiface only (the right face region was blanked out), or a stimulus showing the right hemiface only. If each half of a face was processed independently, accuracy for complete faces should equal the union of the observed accuracy for left and right hemifaces. For upright faces, accuracy exceeded this independence prediction for same-half whole faces (facilitation) and fell below it for different-half whole faces (interference), indicating strong interactive processing. For inverted faces, the interference effect was absent and the facilitation effect was much reduced.

Two papers have taken the approach of disrupting local information to varying degrees and showing that this disrupts identification less for upright whole faces than for other stimuli; such results argue that holistic/configural processing for faces is truly more than the sum of the parts. McKone, Martini, and Nakayama (2001) examined *categorical perception in noise*. After subjects learned two end point faces, they found categorical perception across the identity boundary in a series of intermediate morphs (better discrimination of morph pairs crossing the category boundary than of equidistant pairs drawn from the same side of the category boundary). Heavy noise was then added to the stimuli, with the idea that this would damage the reliability of information from any given local region of the face (e.g., the information extracted from the corner of the left eye might be quite different from one trial to the next). With the noise, categorical perception remained strong for upright intact faces, but was absent for inverted faces and for a single isolated feature (the nose alone).

Using similar logic, McKone (2004) introduced a *peripheral inversion task*. With increasing eccentricity, information from independent consideration of parts should degrade more rapidly than information from holistic/configural processing. For whole faces, as distance from fixation was increased, identification accuracy declined more rapidly for inverted faces (parts only) than for upright faces (holistic plus parts; McKone, 2004; McKone, Brewer, MacPherson, Rhodes, & Hayward, 2007). In contrast, for a single isolated face part (the nose; McKone, 2004) and for objects (dachshund dogs; McKone, Brewer, et al., 2007), upright and inverted performance declined at the same rate, consistent with both upright and inverted versions of the stimulus being processed only as parts.

One way of looking at the results of the categorical perception in noise and peripheral identification techniques is that, as perceptual processing is put under stress (e.g., requiring very fine discriminations, adding noise, etc.), configural processing survives more stress than part-based processing. This suggests that configural processing is more sensitive in some way; that is, it can operate on the basis of less reliable information, or less information, available from the stimulus.

In a direct demonstration of this, Martini, McKone, and Nakayama (2006) reported a saliency bias toward upright in overlaid faces. An upright version of a face was superimposed on an inverted version of the same face, and subjects adjusted the relative contrasts of the two faces until both appeared equally salient. The resulting physical stimulus contained much lower contrast in the upright face than in the inverted face. The saliency bias toward upright was found only for intact whole faces (even when lit from below); scrambled faces produced no saliency bias.

The Strength of the Evidence

To my mind, the evidence from the multiple paradigms presented previously is overwhelming: Upright faces are processed in a manner that inverted faces and nonface objects are not. The holistic/configural effects occur on directly perceptual tasks—as in, for example, the composite effect, the saliency bias, and categorical perception in noise; they also occur in tasks containing a memory component (e.g., part–whole effect). The effects occur for familiar faces (e.g., in the composite effect for famous individuals); they also occur for unfamiliar faces, as in the composite effect for novel faces, the saliency bias, and the disproportionate inversion effect on memory. The differences between upright and inverted faces occur despite the fact that these stimuli are matched in all low-level aspects, such as spatial frequency components, presence of boundaries, brightness, and so on.

When faces and objects were compared, all studies cited used identical tasks for the two stimulus classes. With the exception of Busey and Vanderkolk (2005), all required within-class discrimination (specifically, individual exemplar level discrimination). The different effects for faces and objects cannot be attributed to symmetry differences. Front views of faces are symmetric while the other object stimuli tested so far have not been, but holistic/configural effects also occur for asymmetric views of faces, such as the profile (McKone, 2008).

Also, the different effects for faces and objects cannot be attributed to baseline differences. Even though it is common for upright performance for faces to be better than upright performance for objects, many studies allow direct comparison of the size of effects (inversion, composite, etc.) because faces and objects were matched in the inverted orientation (e.g., Robbins & McKone, 2007, Experiment 1). Also, overall levels of performance seem not to matter as long as ceiling or floor effects are avoided (e.g., identification of upright objects can be easier than or equal to identifying upright faces and yet faces still show larger inversion effects; Robbins, 2005; Robbins & McKone, 2007, Experiment 3; Yin, 1969, Experiment 3).

Cases Where Faces Were Not Processed Configurally

Evidence such as that reviewed before has led to a general consensus by face-recognition researchers that faces are processed holistically/configurally, although a number of cases in the literature have not supported this conclusion. However, in the cases that I am aware of, there is almost always a simple explanation of the failure to find holistic/configural processing. Most commonly, the stimuli were of quite artificial appearance, rather than real faces. These include unnatural schematic drawings (Hannigan & Reinitz, 2000; see discussion in McKone & Peh, 2006; also see Martelli, Majaj, & Pelli, 2005; Schwarzer, 2002) and early Identikit faces (Bradshaw & Wallace, 1971). It is perhaps not surprising that artificial "faces" are processed like objects rather than like faces.

Another problem arises when attention-attracting hairstyles are used. These can provide cues to memory that outweigh the real face information (e.g., see Duchaine & Weidenfeld, 2003). I suspect this was a factor in the Loftus et al. (2004) finding, from state-trace plots, that only a single internal variable was required to describe inversion effects for faces versus houses when the faces were computer-generated rather than natural images. Given that the faces appeared somewhat unnatural in feature shape and placement and also had quite unusual hair, this single internal variable was presumably reflecting a reliance purely on part-based processing.

Occasionally, claims have been made that apparent holistic/configural processing can be attributed to decision-level effects. Wenger and Ingvalson (2002) used an unusual variant of the part-whole paradigm and showed that, in a task where subjects were required to respond "same" or "different" to two features of the face successively, apparently interactive processing between the features could be partially attributed to the response made to one feature biasing the response made to the other feature. For example, if subjects said "same" to one feature, they were likely to say "same" to the other. This seems unsurprising, but provides no evidence that holistic/configural effects in general can be attributed to decision-level effects rather than perceptual processing.

Gauthier and Bukach (2007) attempted to attribute the aligned–unaligned composite difference to a decision bias toward responding "same" in the aligned condition. However, McKone and Robbins (2007) pointed out that there is no reason why any response bias should differ between aligned and unaligned conditions when these are randomly intermixed (as in most experiments), and that the composite effect occurs not only in same–different tasks but also in naming tasks, where the issue of response bias does not arise. Finally, proponents of the decision-level idea have never put forward an explanation of how decision biases would produce differences between faces and objects in tasks with equivalent decision requirements for both stimulus types.

Consistent Evidence From Neuroimaging and Neuropsychology

The focus of the present chapter is on differences in style of computational processing, as revealed in behavioral studies. It is worthwhile briefly noting, however, that faces and objects are also processed differently at the neural level and that links have been demonstrated between these neural differences and holistic/configural processing. (For more extensive reviews, see Kanwisher & Yovel, 2006, for neuroimaging and Duchaine, Yovel, Butterworth, & Nakayama, 2006, for prosopagnosia.)

The evidence of face-specific cortical processing comes from three sources. First, in neuropsychological cases (both acquired and developmental), there is a double dissociation between prosopagnosia and object agnosia. Prosopagnosics exist who have extremely poor recognition of faces in combination with perfectly normal within-class discrimination of objects (e.g., Duchaine, Dingle, Butterworth, & Nakayama, 2004; McNeil & Warrington, 1993; Sergent & Signoret, 1992). A few cases have also been reported of the reverse pattern (e.g., Assal, Favre, & Anderes, 1984). Most famous is CK, who was severely object agnosic but could recognize faces at normal or above-normal levels, even in very difficult formats (e.g., Mooney faces, overlaid cartoons of multiple individuals; Moscovitch, Winocur, & Behrmann, 1997).

Second, neuroimaging studies using fMRI have revealed a face selective area known as the fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997) in the fusiform gyrus that responds two to three times more strongly to within-class discrimination of faces than to within-class discrimination of other objects (e.g., flowers, hands, birds, cars; Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher et al., 1997). In contrast, other areas of extrastriate cortex respond more strongly to objects than to faces (e.g., Lateral Occipital Complex; see Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006).[1]

The final source of evidence comes from monkey single-cell recording studies. It has been known for a long time that monkey inferotemporal cortex contains face-selective cells (e.g., Perrett et al., 1985). Recently, Tsao, Freiwald, Totell, and Livingstone (2006) found a dense cluster of such cells. Starting from fMRI scans and using the same faces-versus-object localizer that is usually used on humans, they located a face-selective region labeled the "middle face patch." Recording from more than 100 single cells in this patch, they found 97% of visually responsive neurons were strongly face selective in comparison to a wide range of objects, including bodies and hands.

Returning to configural processing, some evidence links the human face-specific cortical areas to the core behavioral effects. In neuroimaging studies, fMR-adaptation procedures have shown that the FFA demonstrates an inversion effect on discrimination of individual faces: The BOLD reduction from repeating a face is strong for upright faces but weak or absent for inverted faces (Mazard, Schiltz, & Rossion, 2006; Yovel & Kanwisher, 2005).

More directly, Schiltz and Rossion (2006) implemented a version of the composite effect, again using fMR-adaptation to examine BOLD response in the FFA. Subjects made judgments to top halves of faces and were instructed to ignore the bottom halves; the bottom halves were either all the same (in some blocks) or all different (in others). Across each block, activation in the different-bottom condition decreased less than in the same-bottom condition, thus arguing that the FFA was integrating the top and bottom halves into new wholes. The effect occurred only when the faces were upright rather than when they were inverted; this argues that it did not merely reflect a general inability to restrict attention to the top half.

Neuropsychological evidence is also consistent with the idea that face-specific processing areas perform holistic processing. Prosopagnosics usually show weak or no inversion effects for faces (Duchaine & Nakayama, 2006), consistent with their (poor) recognition being driven by part-based processing, even for upright faces. Some even show a reversed inversion effect (Farah, Wilson, Drain, & Tanaka, 1995), suggesting that a malfunctioning holistic system can grab upright faces and suppress the part-based processing that would otherwise occur in this orientation. In further support, the opposite case of the antiprosopagnosic CK shows much larger inversion effects than controls (Moscovitch & Moscovitch, 2000).

THEORETICAL IDEAS ABOUT HOLISTIC/ CONFIGURAL AND PART-BASED PROCESSING

The empirical evidence shows that faces are recognized through computational procedures different from those used for objects. It is also clear that the theoretical difference must have something to do with local components being processed relatively independently of each other in inverted faces and objects, but being processed in a strongly dependent way in upright faces. What can we say beyond this, however? Can we be more exact about what holistic/configural processing is (and is not)?

Until fairly recently, I have been of the opinion that, really, we had little idea how to conceptualize holistic/configural processing. I have thus tended to stick to an operational definition, focusing closely on the results in the core paradigms. With the array of evidence now available, however, I think the concept of configural face processing can be fleshed out at least somewhat.

Does Holistic/Configural Processing Have Subcomponents?

A first question is whether subcomponents of configural/holistic processing exist. Maurer et al. (2002) proposed that they do. They used "configural" as an overarching term and proposed that this consisted of three subcomponents. They associated each subcomponent with particular core tasks.

The first proposed subcomponent was *sensitivity to first-order relations* (i.e., two eyes above a nose above a mouth). This was proposed to be tapped by *face detection* tasks (i.e., tasks that merely require determining a face is present, rather than identifying it). I agree with Maurer et al. (2002) that there is good evidence that detection can occur independently of identification. Prosopagnosics usually report that they can tell that a face is present and they can see the individual parts, but they just cannot make the face hang together as a person. Also, MEG and reaction time studies in normals indicate that face detection occurs earlier than face identification (Grill-Spector & Kanwisher, 2005; Liu, Harris, & Kanwisher, 2002), as might be expected.

However, some evidence (Liu et al., 2002) indicates that this detection ability relies on independent face parts (i.e., even in scrambled order) rather than relying on having a normal face configuration. This raises the possibility that, rather than coming from the holistic/configural processing stream, face detection could reflect the output of part-based analysis. Further, even if it is the case that face detection can proceed on the basis of first-order configuration, this is not logically sufficient to conclude that face identification does not also refer to first-order structure.

As evidence that it did not, Maurer et al. (2002) argued that the FFA performed face detection (e.g., activity is stronger when the background encourages perception of the stimulus as a face rather than as a vase; Hasson, Hendler, Bashat, & Malach, 2001), but did not perform face identification (no sensitivity to repetition) and was only weakly or not at all sensitive to inversion. The problem with this argument is that subsequent evidence has disproved both of the latter claims. Using the newer technique of fMR adaptation, several studies have shown that

the FFA codes individual identity and that it shows strong inversion effects on identification (e.g., Mazard et al., 2006; Yovel & Kanwisher, 2005).

Maurer and colleagues' (2002) second and third proposed components were *holistic processing*—defined as gluing the features together into a gestalt—and sensitivity to *second-order relations*. Tasks proposed to tap holistic processing were the part–whole effect and the composite effect. Tasks proposed to tap second-order relations were those testing sensitivity to spacing between features (but not local feature shape). As evidence of separability, two factors were proposed to dissociate holistic processing and second-order relations.

The first claimed dissociation was in patterns of childhood development, with holistic processing proposed to be quantitatively mature early (by 6 years of age) and sensitivity to second-order relations proposed to mature much later (10+ years). I am not convinced by either half of this claim. Regarding holistic processing, Maurer et al. (2002) noted two studies that reported part–whole and composite effects no smaller in young children (6 years) than in adults (Carey & Diamond, 1994; Tanaka et al., 1998); a similar result has more recently been found in 4-year-olds (de Heering et al., 2007).

However, none of these studies matched baseline performance across age groups, leading to logical problems in making quantitative comparisons across ages. Indeed, two of these tests (Carey & Diamond, 1994; de Heering et al., 2007) produced the counterintuitive result of *larger* composite effects in 4-, 5-, and 6-year-olds than in adults. This is probably attributable simply to performance in the adult groups approaching a ceiling, but it highlights the point that, in the absence of matched baselines, it is not really possible to know whether children show effects of the same size as those of adults.

Regarding second-order relations, Maurer et al. (2002) noted two studies suggesting that sensitivity to spacing changes was very poor in young children and reached adult levels several years later than sensitivity to local feature changes (Freire & Lee, 2001; Mondloch et al., 2002). This result was also replicated in a later study (Mondloch et al., 2004). However, all these studies used stimuli that failed to match the perceptibility of the spacing and feature changes. Adults found the feature task easier than the spacing task. Thus, the results could simply indicate that development in a harder task lags behind development in an easier task.

Other results indicate no spacing-specific delay. McKone and Boyer (2006) matched spacing and featural changes for effects on perception in adults and then found that even 4-year-olds were as sensitive to spacing as to featural changes. Although they failed to match spacing and feature changes, Gilchrist and McKone (2003) instead matched baseline performance in the unaltered condition across age groups (by using a memory task with smaller learning set sizes for the younger children). Under these circumstances, 6- and 7-year-olds showed as strong a sensitivity to spacing changes as did adults.

The other claimed dissociation between holistic processing and second-order relations was based on the effects of using photographic negatives (contrast reversal). The proposal was that negation affects sensitivity to spacing changes but does not affect holistic integration. Kemp, McManus, and Piggott (1990) found that detection of spacing changes was substantially poorer in negative contrast faces than in positive contrast faces. The other half of the dissociation, regarding holistic processing, was based on Hole, George, and Dunsmore's (1999) finding that a version of the composite effect was as strong for reversed contrast grayscale faces as for normal-contrast faces.

Unfortunately, however, Hole et al. (1999) did not test the usual full composite design—namely, aligned and unaligned versions for upright and inverted faces. Instead, they tested only aligned composites and relied on the difference between upright and inverted conditions being in the opposite direction to usual (i.e., inverted was better than upright) to argue that holistic interference must have occurred for upright faces. To understand why this procedure is a problem in a contrast reversal study, consider that, in fact, the total inversion effect on naming the target half has two components: the (reverse direction) inversion effect arising from holistic interference slowing reaction times for upright but not inverted faces (i.e., the true composite effect, component A) and the (normal direction) inversion effect arising from part-based processing of the individual target half (component B).

For illustration, consider a case where the total inversion effect was –40 ms (i.e., reaction times were 40 ms slower for upright than inverted faces). Presume that, for normal-contrast faces, this is made up of a –100-ms inversion effect on holistic interference and a +60-ms inversion effect on part-based processing (i.e., A = –100 ms, B = +60 ms, total = –40 ms). In interpreting their finding of equal total inversion effects for normal-contrast and contrast-reversed faces as evidence of equal holistic processing, Hole et al. (1999) implicitly assumed that contrast reversal had no influence on inversion effects on *either* holistic interference *or* part-based processing.

There is no guarantee, however, that this is the case. A –40-ms total effect could be made up, for example, of a –70-ms inversion effect on holistic interference and a +30-ms inversion effect on part-based processing (i.e., A = –70 ms, B = +30 ms, total = –40 ms). If this pattern occurred for contrast-reversed faces, then the total inversion effect measure could fail to reveal the presence of weaker holistic processing for contrast-reversed faces than for normal-contrast faces.

Overall, I see no convincing evidence of dissociations between holistic and second-order relational tasks. There are also other good reasons to prefer the more parsimonious idea that first-order structure, second-order relations, and holistic gluing are all aspects of a single form of representation. Maurer and colleagues' proposed subcomponents (2002) associate rather than dissociate in two key ways. Large inversion effects are present on the tasks associated with all three (for holistic and second-order relational tasks, see earlier section; for face detection, think of the difficulty of seeing a Mooney face upside down).

Also, an apparent critical period in infancy applies to both the second-order relational and holistic subcomponents. Le Grand et al. (2001) tested people born with congenital cataracts that allowed no form vision until removal at 2–6 months of age. At 9–21 years of age, these patients had very poor sensitivity to spacing changes (which arose specifically with early visual deprivation to the right hemisphere; Le Grand, Mondloch, Maurer, & Brent, 2003) and a lack of composite effect (Le Grand et al., 2004).

There are also more theoretical reasons to prefer a single form of representation. The problem with associating certain subcomponents with certain experimental tasks is that, when new tasks come along, it can be difficult to slot them into the existing scheme. For example, where does the McKone et al. (2001) categorical perception in noise results fit? Maurer et al. (2002) described these results under the heading of sensitivity to second-order relations, saying that the lack of categorical perception with isolated features or inverted faces arose "presumably because second-order relational information was not available" (p. 257).

But why should the effects be attributed to this component? Showing a face inverted or the nose alone also destroys the first-order arrangement of the face and/or the potential for holistic integration. Why not assign the effect to one of these components? Similarly, to which subcomponent would the saliency bias effect (Martini et al., 2006) be attributed? Also, which subcomponent is responsible for the part-in-spacing-altered-whole effect (Tanaka & Sengco, 1997)? In this case, the manipulation is one of spacing, which would suggest second-order relations, but the method is a variant of the part–whole paradigm, which would suggest the holistic component. Overall, my point here is that the rationale for the association of particular tasks with proposed subcomponents in Maurer and colleagues' theory (2002) is not sufficiently spelled out to make it a useful theory in light of more recent evidence.

A Very Different Theory: Configural/Holistic Processing Is Not Based on Decomposition Into Eyes, Nose, and Mouth

Although it is possible to imagine many different alternative positions to that of Maurer et al. (2002), one alternative worth noting is the theory of Tanaka and Farah (1993). Maurer and colleagues' idea of a special spacing subcomponent to configural processing that is different from sensitivity to other sorts of deviations from the average template (e.g., in individual feature shape) seems implicitly based on the idea that decomposition into named-level parts provides a direct input into the formation of a configural representation.

In contrast, Tanaka and Farah (1993) suggested that the whole-face processing for (upright) faces did not decompose faces into such parts at all. In their terminology, this was *holistic processing* (with no subcomponents). Although I find their original evidence for this idea—which was merely the observation of a part–whole effect—unconvincing, more recent evidence is quite strongly suggestive of it.

This relevant evidence is that configural/holistic processing for faces can operate in the complete absence of part-based processing. Importantly, this is not to say that faces *cannot* be decomposed into parts; clearly, they can (e.g., we can describe the color of someone's eyes or the shape of his or her nose). Instead, the idea is that two independent processing routes exist that can contribute to performance on face-recognition tasks and that these branch off directly from some quite early stage of visual processing (Figure 10.3; Moscovitch et al., 1997; also see McKone, 2004; McKone, Martini, & Nakayama, 2003).

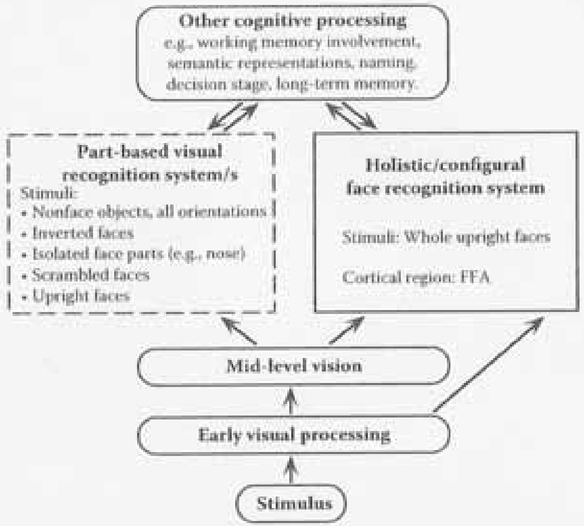

Figure 10.3 A possible neural/functional architecture in which holistic/configural processing is not derived from part-based processing as a preliminary step. This would explain the ability of holistic/configural processing to operate in the absence of part-based processing for faces.

One route is responsible for forming a configural/holistic representation of the whole face. The other, presumably open to nonface objects as well, is responsible for part decomposition. Note that, in this theory, there is no arrow from the part-based box to the configural/holistic box; that is, configural/holistic processing does not derive from part-based processing.

The evidence that configural processing can operate without part-based processing is as follows. Patient CK, who had prosopagnosia without object agnosia, was good at recognizing faces only when they were intact whole faces in the upright orientation. As soon as the normal configuration was disrupted, by inverting or exploding the face or by showing isolated face parts, CK's performance was extremely poor—many standard deviations below that of controls (Moscovitch et al., 1997). Thus, excellent recognition of whole faces occurred at the same time as extremely poor part-based processing.

Similar results have been demonstrated in subjects without brain injury. McKone et al. (2001) found categorical perception phenomenon in noise for upright whole faces, despite its complete absence for the nose, which was the most discriminating feature between the particular faces tested, as well as its complete absence for inverted faces. McKone (2004) also reported that a particular high-contrast "Mooney face" (Mooney, 1957; Figure 10.4) was, for most people, perceivable as a face *only* in the upright orientation. Approximately 80% of people (a number

Upright Inverted

Figure 10.4 A Mooney face (Mooney, 1957, *Canadian Journal of Psychology, 11*, 219–226) that is particularly difficult to see. Approximately 80% of people can see it upright, but not inverted—no matter how often the stimulus is viewed. The face is a young, attractive Caucasian woman lit from top-right. Rotation in the image plane causes the face percept to drop out somewhere around 90° for most people (range = 45–135° across individuals). (After McKone, E., 2004, *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 181–197.)

based on formal data in McKone, 2004, and on hands-up tests during conference presentations) can see this particular face when it is upright but do not see it at all when the stimulus is inverted. For these people, the regions of the stimulus that form the face features are only perceived as face components when the whole face is perceived: that is, the parts do not exist without the whole.

Results from a study on learning to differentiate between identical twins (Robbins & McKone, 2003) are also consistent with holistic/configural processing operating without part-based processing. For subjects who saw a set of twins inverted, the only local feature that was sufficient for learning the twins was a minor difference in combing of the eyebrows. All subjects who successfully identified the twins reported using this strategy, and the subjects who did not report this strategy failed to identify the twins. Of subjects shown upright stimuli, however, no person reported noticing the eyebrow difference, yet all learned well and also showed an aligned–unaligned composite effect for the twins. Thus, good identification and holistic processing were possible in the apparent absence of awareness of even the most useful single local cue that could assist identification.

The Current McKone View of Configural/ Holistic and Part-Based Processing

Based on the evidence available at the time of writing, my view of the nature of configural/holistic processing is along the following lines. I think it references a detailed coding of face structure. This includes first-order information about basic face structure *and* second-order information about distances between regions of the face, *and* second-order information about exact feature shape, *and* exact face shape aspects that are less easily labeled as features (e.g., structure of cheekbones, depth of eye sockets, and angle of forehead in a profile view). Also, it intrinsically glues all these together—perhaps because they were never separated in the first place.

I also find that a useful analogy is one suggested to me by Paolo Martini and Ken Nakayama, in terms of the size of receptive fields: Although part-based processing can be thought of as applying "lots of little receptive fields" to a face or object and then summing the results, configural/holistic processing can be thought of as "one big receptive field" applied to the entire face region. (The reason for the quotation marks here is that the term "receptive field" normally refers to a region on the retina, while face and object recognition are largely size- and position invariant and the relevant cortical areas are nonretinotopic.)

Finally, an important aspect of configural/holistic processing is that it is more sensitive than part-based processing in that it can operate on the basis of less information in the stimulus. Theoretically, this might be a result of having a larger receptive field. Functionally, it may serve to enhance the perceptual salience of faces and to attract attention to them in cluttered visual settings.

It is also worthwhile to make a quick comment on what I think configural/holistic processing is *not*. Occasionally, one runs across the idea that the term "configural" implies that the representation contains only low spatial frequency information (noting that spatial frequency for faces usually refers to cycles per face width, rather than absolute spatial frequency). I see no evidence for this. All the evidence suggests that configural/holistic processing is the factor driving humans' remarkably good discrimination of so many different individual faces.

Sergent (1984) showed that low spatial frequency information, taken alone, is suitable for certain limited types of face discrimination (e.g., men from women, fat faces from thin faces), but is not sufficient for distinguishing between individuals of very similar general appearance. Indeed, in discriminating identical twins—the ultimately demanding face recognition task—it would surely be of value to consider information at all spatial scales. More directly, Goffaux and Rossion (2006) reported significant part–whole and composite effects for faces filtered to contain only high spatial frequency components, although the effects were larger for medium spatial frequency images and larger again for low spatial frequency images. Overall, low spatial frequency information may drive holistic/configural processing most strongly, although high spatial frequency detail is represented as well (also see McKone et al., 2001, p. 595).

What about part-based processing? Traditionally, face-recognition researchers have tended to focus on configural/holistic processing because it is special to faces, and they have tended to say little about what they understand part-based processing (for faces or objects) to mean. I continue that tradition here. I have little expert knowledge about the processes of object recognition, beyond knowing that the empirical literature shows differences between faces and objects. By the term "part-based processing," I merely mean some form of image decomposition after which the resulting components are treated in a relatively independent manner.

I take no particular position on whether this part decomposition is into abstract three-dimensional view-invariant components such as "geons" (Biederman, 1987) or whether it involves decomposition into much smaller image sections, as in the chorus of fragments theory (Edelman & Intrator, 2000), and so on. I also have no particular position on how context effects arise for objects (e.g., object superiority effects) or how global/local effects arise (e.g., Navon figures), beyond suggesting

that these must have some explanation other than the type of perceptual integration that has been shown to occur for faces.

A Quick Comment on Terminology

In the present chapter, I have used "configural/holistic" to refer to the style of processing special to (upright whole) faces. In previous research articles, I have deliberately alternated "configural" and "holistic" across successive papers because I have seen no reason to discriminate between them in meaning. Other face-recognition researchers have used terminology differently. The Maurer et al. (2002) usage was discussed earlier: "configural" as an overarching term with "holistic" and "second-order relational" as subcomponents. Other authors use "configural" quite differently from Maurer and colleagues' use; it is quite common to see configural information closely equated with the empirical manipulation of changing spacing between features (Peterson & Rhodes, 2003; Yovel & Kanwisher, 2004). In Tanaka and Farah (1993), "holistic" was used as the general term, with no division into subcomponents. Some authors use "configurational"—again, as a general overarching term (Hole, 1994).

What are the advantages and disadvantages of the different terms? Unfortunately, all terms are far from perfect. "Holistic" captures the idea of strong perceptual integration but fails to acknowledge that orientation matters. "Configural" can be more useful in getting across the idea that orientation is important because configuration, in the sense of the arrangement of parts, is different in upright and inverted orientations. On the downside, configuration is sometimes taken by researchers to mean distances between "feature blobs," or low spatial frequency components, losing the relevance of much of the information in the face.

Of course, both terms are problematic in that they are also widely used outside the face perception literature, with different meanings (e.g., "holistic" for any processing of global structure that occurs in context and gestalt effects, "configural" to describe any sensitivity to even first-order arrangements of parts, and "holistic" for attentional field dependence). In the absence of any better terminology, I will stick with "holistic/configural" for the rest of this chapter.

FACES, OBJECTS, AND THE EXPERTISE HYPOTHESIS

In the previous sections, I described the evidence that, normally, faces are special in comparison to objects in that only faces receive holistic processing. In general, this claim is uncontroversial. (When it has been questioned, authors have appeared unaware of the key findings that were reviewed in the previous section.) Far more controversial, however, has been the question of whether faces are always special—suggesting that it is their structural form that is essential for holistic processing—or whether objects might also be processed holistically under one specific circumstance, such as in the case where a person has become an expert with a certain type of object class (e.g., a dog-show judge or a car expert). People who are expert in within-class discrimination of objects are rare, but the case is still theoretically important in that if such people show holistic processing for their objects of expertise, then this would demonstrate that holistic processing is potentially applicable to a wide range of structural forms.

Although the expertise hypothesis makes testable predictions about objects of expertise, it is fundamentally a theory about the origin of special processing for faces. The theory was first put forward by Diamond and Carey (1986; also see Carey's developmental work, e.g., Carey, 1992), who proposed that the reason why upright faces are processed holistically is that, by the time a person is an adult, he or she has had many years of experience in individuating upright faces and has become functionally expert at doing so. In contrast, inverted faces are not processed holistically because these stimuli are much rarer; adults have had little experience with them and remain functionally poor at telling them apart. Similarly, the reason for the lack of holistic processing for nonface objects (in both upright and inverted orientations) is a lack of expertise in individual level discrimination (e.g., telling Labrador 1 apart from Labrador 2).

This lack of experience arises not because appropriate visual input is not available (we may see many upright Labradors in our lifetime) but rather because there is little motivation or functional need for individuation. For faces, individuation is critical to appropriate social behavior. In most everyday cases of object recognition, however, it is sufficient to discriminate at the between-class level. We need to discriminate trees from tables but, unless we are foresters, most of us do not need to tell one tree from another.

Of course, even for ordinary people, in some cases it is important to recognize one or two individuals of an object class, such as *my* dog or *my* toothbrush. However, recognition remains poor for exemplars of the object class beyond the one or two highly familiar ones, and even the familiar items are often recognized largely by some obvious single feature (e.g., family members buy toothbrushes in different colors) or from the context (e.g., the dog in your backyard is probably your dog).

Predictions of the Expertise Hypothesis

The expertise hypothesis has been used to draw a number of predictions. The most straightforward is that objects of expertise should be processed like faces, in that both should receive holistic processing in their familiar upright orientation.

The second involves the development of holistic processing for upright faces in children. If, as Carey has suggested (e.g., 1992), face expertise takes many years of experience to achieve, then holistic processing for upright faces and corresponding differences between upright and inverted faces might emerge quite late in development. In terms of the amount of experience required to develop expertise, studies of other types of expertise (e.g., in chess or music) have indicated that 10 years or so of intensive practice is required (e.g., Gobet & Simon, 1998).

Probably partly for this reason, early developmental studies focused on the idea that there might be an "encoding switch" from part-based processing for faces to holistic processing for faces at around 10 years of age (Carey, Diamond, & Woods, 1980). In fact, the expertise hypothesis per se makes no specific prediction about the age at which holistic processing should emerge. It merely predicts that this should not occur until the child has become a "face expert"; for all we know, this could occur at 15 years, or 3 years, or as an infant.

It does, however, at least place a strong limit on the amount of experience that should be required to produce holistic processing for objects in objects-of-expertise studies. Specifically, the amount of experience required to produce holistic processing for objects should be the same as the amount of experience required to produce it for faces. The expertise hypothesis proposes only that practice is the causal mechanism. It includes no proposal that the effects of practice would depend on stage of development.

Diamond & Carey (1986, pp. 116–117) argued that a certain amount of experience with dogs, gained largely as an adult, corresponded to a similar amount of experience obtained with faces, beginning as an infant, in its ability to produce holistic processing. The same assumption that developmental stage is not critical is made by multiple subsequent tests of the expertise hypothesis (e.g., Busey & Vanderkolk, 2005; Gauthier & Tarr, 1997; Xu, Liu, & Kanwisher, 2005) in which subjects gained expertise largely or entirely as an adult. Thus, the expertise hypothesis predicts that if, say, children showed configural processing at 6 years of age, then 6 years' experience should be sufficient for holistic processing to emerge with dogs (possibly adjusted somewhat for the fact that even dog freaks probably see fewer dogs than faces).

A third prediction of the expertise hypothesis again derives from the proposed explanation of inversion effects for faces. According to the expertise hypothesis, it is the greater experience with upright faces that leads to holistic processing for upright (common format) but not inverted (rare format) faces. This then predicts that other methods of varying the natural frequency of different face formats should also affect holistic processing. One such method is rotating a face in depth; front-on views are far more common than profile views. Thus, the expertise hypothesis predicts that holistic processing should be weaker for profiles than for front views.

The first two of these predictions have been the focus of substantial empirical investigation. In contrast, the third prediction has only recently been addressed. In the remainder of this section, I concentrate on evaluating predictions one and two; prediction three is left until a later section.

Initial Evidence for the Expertise Hypothesis

The expertise hypothesis has long held sway in the literature and, indeed, had reached the status of being the zeitgeist in 2007. This is partly because it offers a coherent theoretical proposal about why faces might be special, but it is also because initial evidence appeared to provide compelling support for it. In fact, as we will see in the next section, all of this early evidence has since been refuted, but these early studies are described because they still have a powerful influence in the field. Even today it is not uncommon to see authors citing the initial papers without the corrections provided by subsequent literature.

One component of the early evidence came from studies of children's face recognition. Carey et al. (1980) reported that children's memory for faces was not affected by upright versus inverted orientation at 6 years of age. However, an inversion effect emerged at 10 years, so they therefore argued that holistic processing for upright faces emerged somewhere between these two ages.

The other component comes from Diamond and Carey's (1986) classic study of dog experts. This study tested the expertise hypothesis prediction that inversion effects on memory for dogs in dog experts should be larger than for dogs in novices and, indeed, might be similar to the size of the inversion effect for faces. One experiment showed some suggestion of this pattern, but the expertise × inversion interaction was not significant. In the follow-up experiment that became famous, mean expertise was increased to 31 years, and the breeds of the dog stimuli were carefully matched to the breeds of expertise of the show judges (this had not been done in the first experiment). A striking expertise effect was then apparent. Dog novices showed, as usual, a small inversion effect for dogs. Dog experts showed a very large inversion effect that was as large as that obtained for faces.[2]

Subsequent Falsification of the Expertise Hypothesis

Regarding the early results in children's development of face processing, there were clear problems with the initial studies. Carey and colleagues' (1980) inversion effect study suffered from floor effects in the 6-year-old group, leaving little room for inverted to be poorer than upright. Subsequent studies without floor effects have reported inversion effects in 6-year-olds (e.g., Carey, 1981), 4- and 5-year-olds (e.g., Pellicano et al., 2006), and 3-year-olds (Sangrigoli & de Schonen, 2004). In the one study that equated baseline across age groups (Carey, 1981), the size of the inversion effect was the same in young children as in adults. Using looking time and habituation paradigms, inversion effects have also been found in infants (e.g., Bhatt, Bertin, Hayden, & Reed, 2005; Cohen & Cashon, 2001).

More direct means of testing holistic processing have confirmed the implication from the inversion findings that holistic/configural processing is present in young children. The part–whole effect has been obtained in 6-year-olds (Tanaka et al., 1998) and 4- and 5-year-olds (Pellicano & Rhodes, 2003). The part-in-spacing-altered-whole effect has been obtained in 4- and 5-year-olds (Pellicano et al., 2006). The composite effect has been obtained in 6-year-olds (Carey & Diamond, 1994) and 4- and 5-year-olds (de Heering et al., 2007); in both cases, the effect was numerically larger than in adults. Sensitivity to exact distances between face parts has been obtained in 6-year-olds (Gilchrist & McKone, 2003; Mondloch et al., 2002) and 4-year-olds (McKone & Boyer, 2006). In infants, Cohen and Cashon (2001) found that a new face composed of old parts was treated as new rather than old.

In summary, data from children clearly indicate that holistic processing is present and, indeed, strong by 4 years at the latest. No data are available on children from the 1- to 3-year age range, but there is certainly some suggestion that holistic processing is present even in infancy.

Turning to objects of expertise, studies subsequent to Diamond and Carey's (1986) original dog expert studies have almost universally failed to find evidence suggesting holistic processing in experts. Regarding the basic inversion effect, no study has replicated Diamond and Carey's finding of face-sized inversion effects for objects of expertise. All relevant studies of which I am aware (as of August, 2007) are reviewed in Table 10.1. As can be seen, the general finding is better described as somewhere between no increase and a small increase in inversion effects with expertise. Importantly, this includes even in our own study that directly replicated Diamond and Carey's original design, employing dog experts looking at side-on photographs of their breed of expertise.

A likely explanation of Diamond and Carey's (1986) original result (see Robbins & McKone, 2007) is that dog experts (American Kennel Club judges) were preexperimentally familiar with the particular dogs used as stimuli (taken from AKC training manuals), along with their names. This would provide an artificial boost to memory in the upright orientation because this is the orientation in which the experts would previously have seen the stimulus dogs (e.g., in training manuals) and because having access to a name to remember as well as a picture is known to improve memory (e.g., Paivio, 1986).

It is also crucial here to reiterate the logic of the interpretation of inversion effects. Even if inversion effects do become slightly stronger with expertise, nothing in this finding per se requires that the increase must have come about because experts are learning to use configural/holistic processing in the upright orientation (cf. Valentine, 1991). An increase could arise from experts learning better part-based processing in the upright (most experienced) orientation, as is suggested by Robbins and McKone's (2007) finding of excellent transfer of expertise to contrast-negative dogs. The fact that larger inversion effects were caused by holistic/configural processing for upright faces does not mean that increased inversion effects for objects must also be holistic in origin.

The results of the more direct tests of configural/holistic processing are thus critical. Results from all relevant studies were shown in Tables 10.2, 10.3, and 10.4. The findings are clear-cut. The part–whole effect does not increase with expertise. The part-in-spacing-altered-whole effect does not increase with expertise. The composite effect does not increase with expertise; indeed, Robbins and McKone (2007) found no composite effect at all for dog experts looking at their breed of expertise, despite these experts having very high levels of expertise (a mean of 23 years' experience and the ability to match dogs as accurately as faces). Thus, improvements in performance with expertise and the small increase in inversion effects apparent in some studies must be coming from part-based processing, rather than holistic/configural processing.

Only two results in the literature might appear to challenge this conclusion (Busey & Vanderkolk, 2005; Gauthier, Curran, Curby, & Collins, 2003). In both cases, strong effects of expertise have been found on measures that the authors claimed tap holistic processing. Both used nonstandard tasks. Robbins and McKone (2007, Section 5.3) argued in detail that one of these (Gauthier et al.) definitely does not measure integration of parts into a whole at a perceptual level; instead, it measures merely the inability to ignore competing response cues from notionally irrelevant information (as in the Stroop effect). They also argued that there were reasons to doubt the validity of the other. (Busey and Vanderkolk rely on a model with questionable assumptions that has never been tested on faces.)

Finally, it is worth commenting on levels of expertise. In the face of null findings on the core paradigms, a common reply from proponents of the expertise hypothesis is that, of course, one would not expect the effects to be as big in experts as they are for faces because the level of expertise for objects remains lower than it is for faces. This is where the developmental face data come into play. The early emergence of holistic face processing disposes of the idea that perhaps experts might show face-like processing if only they were "more" expert. If babies and 4-year-old children show clear and statistically significant holistic effects for faces—despite the well known difficulties of testing in this age range (e.g., it is very easy to get no effect because the child did not understand the task), then surely the 10+ years of expertise used in many object expertise studies should also be sufficient for significant effects to show up in the much more reliable case of testing adults.

Relevant Data From Neuroimaging and Neuropsychology

Results relevant to neural substrates also support the idea that face-like processing does not emerge for objects of expertise (for review, see McKone, Kanwisher, & Duchaine, 2007). Findings from neuropsychology are the most dramatic. In cases of brain injury, the expertise hypothesis predicts that ability to recognize objects of expertise should always track ability to recognize faces (e.g., if one is damaged, both should be damaged). In contrast, the idea that faces, as a structural form, are recognized via face-specific cortical areas predicts that objects of expertise should track other objects and dissociate from faces.

Evidence clearly favors face specificity. No cases have been reported in the literature following the expertise hypothesis prediction, but a double dissociation between faces and objects of expertise has been reported. Some patients cannot recognize faces, but retain or gain expertise in individuation of objects: most famously, the farmer, WJ, could recognize his sheep, but not his family (McNeil & Warrington, 1993; also see Duchaine et al., 2004; Sergent & Signoret, 1992). Others show the converse pattern of normal face recognition with impaired recognition of former objects of expertise (e.g., Assal et al., 1984; Moscovitch et al., 1997).

In neuroimaging, seven studies have tested the expertise hypothesis prediction that a BOLD response in the FFA should increase for objects of expertise compared to the same objects in novices. Three reported no change at all in FFA response (Grill-Spector et al., 2004; Op de Beeck et al., 2006; Yue, Tjan, & Biederman, 2006), two reported very small and nonsignificant trends toward an expertise-related increase (Moore, Cohen, & Ranganath, 2006; Rhodes, Byatt, Michie, & Puce, 2004), and two reported significant increases to a level that did not approach the level for faces (Gauthier, Skudlarski, Gore, & Anderson, 2000; Xu et al., 2005). Of the five studies that also examined response in other areas of extrastriate cortex, all five reported larger expertise-related increases outside the FFA than within it (Gauthier et al., 2000; Moore et al., 2006; Op de Beeck et al., 2006; Rhodes et al., 2004; Yue et al., 2006).

Taken together, these results provide no evidence for the special relationship between expertise and the FFA that was predicted by the expertise hypothesis.

One alternative proposal (McKone, Kanwisher, & Duchaine, 2007; Xu, 2005) is that expertise effects arise primarily in the same cortical regions responsible for recognizing the objects in novices and that the small and inconsistent effects in the FFA simply reflect general attention-related increases in blood flow arising from experts being more interested in their objects of expertise than novices. Some effects could also arise from inclusion of nontarget neural material ("partial voluming"): fMRI voxels are quite large and cubic in shape, and their edges are very unlikely to correspond to the boundaries of cortical regions.

VIEWPOINT, RECENT EXPOSURE HISTORY, AND FACES VERSUS OBJECTS

The results reviewed in the previous section indicate that holistic processing is not learned with experience for objects. In contrast, improvements in object recognition with expertise appear to have their origin in improved part-based processing. This suggests a dissociation between holistic/configural and part-based processing in terms of their patterns of sensitivity to prior experience. Other ways of testing the effects of prior experience are discussed in this section. These again show face–object dissociations.

First, consider holistic/configural processing for faces. With rotation from upright to inverted in the image plane, holistic/configural processing falls off in a bell-shaped manner and is absent in the range from approximately 135° of rotation to 180° (Martini et al., 2006; McKone, 2004; McKone et al., 2001). This result could potentially be explained based on differential experience with different rotations, but two findings argue against this interpretation. Experimental practice with inverted faces does not induce any holistic/configural processing: this includes hundreds of trials of practice (McKone, 2004), thousands of trials (Robbins & McKone, 2003) and tens of thousands of trials (McKone et al., 2001). More anecdotally, I have been looking at the Mooney face in Figure 10.4 for five years and have never seen the face inverted.

The other finding is that rotation in depth (front through profile, all faces upright) has no effect on configural processing. Despite the fact that people have substantially more experience with front and three-quarter views than with profiles, McKone (2008) found that the composite effect was equally strong in all views. (Note that this also refutes the third prediction of the expertise hypothesis.)

Now consider object recognition and part-based processing in general. For objects, naming latencies increase linearly with rotation away from the canonical view in the image plane and in depth (e.g., Jolicœur, 1985; Palmer, Rosch, & Chase, 1981). These rotation effects disappear rapidly with practice (e.g., Jolicœur, 1985; Tarr, 1995), usually within 3–30 trials per stimulus. Figure 10.5 illustrates an example where the misorientation curve became flat in less than 20 exposures to each of 54 common objects (McKone & Grenfell, 1999).

People can also learn entirely new object classes easily, even when first exposed to these as adults (for readers 30+ years in age, think of mobile phones; for readers 40+, think of computer mice). Results show that part-based processing for objects

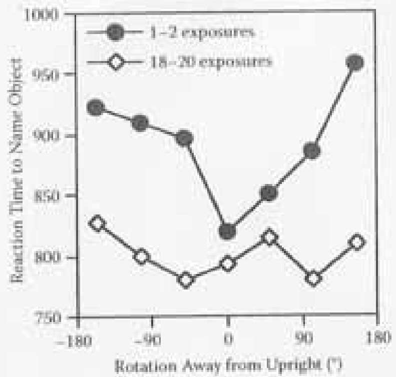

Figure 10.5 In contrast to the findings that holistic/configural processing for inverted faces shows no learning, even with thousands of trials of practice, "upright-like" processing for objects (presumably part-based) is learned very rapidly for inverted and other misoriented objects: Rotation effects disappear in less than 20 exposures to each of 54 common objects. (Results are from McKone, E. & Grenfell, T., 1999, *Perception and Psychophysics*, *61*, 1590–1603.)

is strongly sensitive to experience, including very recent exposure history. The same is true of part-based processing for faces. Despite the lack of view frequency effects on holistic/configural processing, McKone (2008) found an effect of view frequency on part-based processing, as evidenced by a profile decrement that was as strong for disrupted-configuration faces (unaligned, inverted) as for intact-configuration faces (upright, aligned).

Taken together, these results argue that part-based processing for objects (and faces) is strongly sensitive to experience, including recent experience obtained as an adult, and improves easily with practice. Holistic/configural processing for faces, in contrast, seems unaffected by experience: It is always good for upright faces, regardless of depth rotated view, and always poor for inverted faces, regardless of amount of practice.

DOES EXPERIENCE *EVER* MATTER FOR HOLISTIC/CONFIGURAL PROCESSING FOR FACES?

In contrast to the general lack of experience effects referred to previously, in two specific circumstances, holistic/configural processing is sensitive to experience. I have already noted the existence of a critical period in infancy: Cataract patients who are not exposed to faces for the first 2–6 months of life never show a composite effect (Le Grand et al., 2004). Thus, experience with faces in infancy seems to be essential to developing holistic/configural processing.

The second circumstance involves other-race effects. Holistic/configural processing is one factor involved in the poor discrimination of other-race relative to own-race individuals. Both the composite effect and part–whole effects are weaker for other-race faces than for same-race faces (Michel, Caldara, & Rossion, 2006; Michel, Rossion, et al., 2006; Tanaka, Kiefer, & Bukach, 2004). This suggests that holistic/configural processing is affected by experience with particular face subtypes. Further, normal levels of holistic processing can be induced for trained other-race individuals after only 1 hour of practice (McKone, Brewer, et al., 2007).

In summary, it seems that for faces *as an entire class*, experience matters to holistic/configural processing only in that there is a critical period for the activation of holistic processing in early infancy. For specific *subtypes* of faces (own race vs. other race), it may be that holistic processing can be switched on through experience or off through lack of it, and that this can occur quite rapidly, even for an adult. Crucially, however, this latter flexibility applies only if the face is upright.

WHAT IS THE ORIGIN OF SPECIAL PROCESSING FOR FACES?

This chapter has reviewed an extensive set of findings arguing that faces and objects differ critically in style of processing and in patterns of sensitivity of that style to prior experience. The key results can be summarized as follows. Holistic/configural processing is limited to the structural form of upright faces (in any depth viewpoint), cannot be learned for objects, and is insensitive to amount of experience with different views/rotations of faces. It is sensitive to experience only in that exposure to faces in infancy is required to activate it and that it can weaken or strengthen for particular subtypes (e.g., races) of faces. In contrast, part-based processing, as occurs for both faces and objects, is strikingly sensitive to experience. It improves rapidly with experimental practice to the point where misorientation effects can quickly disappear. It also has no critical period in infancy for exposure to particular structural forms.

What is the origin of these differences? As McKone, Kanwisher, and Duchaine (2007) have noted, it is clear that generic expertise is not the origin. Instead, they noted that this leaves researchers with two types of theories about the limitation of holistic/configural processing to the structural form of faces. These differ in whether they include an innate representation of face structure.

Perhaps the most obvious hypothesis is of an *experience-expectant innate template*. This theory proposes that a representation of face structure has developed via evolutionary processes, reflecting the extreme social importance of faces; at the same time, the visual system has maintained an independent and more flexible generic system suitable for recognizing any type of object. Within such a theory, the following components would be necessary to explain the face-recognition data I have reviewed:

- The innate "template" would code at least the basic structure of a face. The form of any such representation is not understood, but it could possibly take the form of eye blobs above nose blob above mouth blob, as in the Morton and Johnson (1991) CONSPEC theory (also see de Haan, Humphreys, & Johnson, 2002).

- The template must provide the developmental impetus not only for good face recognition (as Morton & Johnson, 1991, suggested), but also for the emergence of holistic processing and the grouping of face-selective neurons seen as the FFA in adults; how it would do so remains unknown.
- The activation of the template must rely on appropriate input during a critical period in early infancy, without which it would no longer function.
- Following a normal infancy, the coding of face structure must remain general enough to allow holistic processing to be applied to initially non-experienced subtypes of faces following practice, but it must be permanently tuned to the upright orientation of faces.
- To explain the lack of depth viewpoint effects, the template must be three-dimensional in structure or there must be multiple innate templates, each describing different views.

The experience-expectant innate template theory can explain all the results I have reviewed here, plus all other results of which I am aware. It can explain the existence of developmental prosopagnosia (e.g., Duchaine et al., 2006) and the fact that, anecdotally at least, this appears to be strongly heritable. It can explain the fact that all typically developing humans choose to individuate conspecifics (members of their own species) based on the face, rather than on some other body part. Despite extensive opportunity to develop expertise with, say, hands, adults fail to do so and remain poor at identifying these stimuli compared to faces.

It can also explain a finding that 6-month-old infants can discriminate individual monkey faces, although 9-month-olds and adults have lost this ability (Pascalis, de Haan, & Nelson, 2002). This finding is similar to the perceptual narrowing with lack of experience that occurs during infancy for phonemes of nonexperienced languages, which is usually taken as evidence for an experience-expectant innate coding of all possible phonemes.

An alternative idea is an *infant experience plus other factor* theory. In many ways, this appears to be a viable possibility. It can explain the core finding—that holistic/configural processing is limited to faces in adults—by proposing that any innate special visual ability is for the style of processing rather than the particular structural form and that this becomes tuned to faces due entirely to biased exposure to faces in early infancy, which arises from some factor other than an innate face template. That is, upright faces are the only homogeneous stimuli for which individual-level discrimination is practiced during the critical period for holistic/configural processing in infancy.

Importantly, the theory is not merely another version of the expertise hypothesis. The mechanisms supporting face expertise in the infant brain would necessarily be different from those supporting general object expertise in the adult brain. Without this assumption, it should be possible to learn holistic/configural processing for objects as an adult, but it is not. It is also important to note that this type of theory does not rule out all innate contributions, but merely innate contributions based on a visual representation of face structure. Innate contributions based on, for example, other visual preferences or auditory abilities would be possible.

In terms of explaining other relevant findings, the performance of the theory of infant experience plus other factor is rather mixed. It provides a good explanation of the choice of the face for conspecific individuation: This would arise because infants experience more faces than any other stimuli. It also provides a potential explanation of the heritability of developmental prosopagnosia: This could arise if something is genetically wrong with the "other factor," rather than with a face template; however, note that this requires specifying a reasonable other factor, which is not an easy task (see following comments).

A possible difficulty for the theory, however, is the lack of viewpoint frequency effects on holistic/configural processing and the different tuning patterns for rotations in the image plane versus in depth. The theory can explain the fact that holistic/configural processing occurs for upright but not inverted faces because inverted faces are presumably rare in infancy; however, it would then need to develop some principled explanation of why profile faces, which presumably are also rare in infancy, show configural processing as strong as that for the common front-view face.

Even more importantly, for this theory to be viable, it would be necessary to be able to identify a workable "other factor." McKone, Kanwisher, and Duchaine (2007) considered four possibilities. Three of them have clear difficulties.

Simion, Valenza, Cassia, Turati, and Umilta (2002) have suggested that infants' preference for face stimuli is based on a preference for stimuli with more elements in the upper half of the visual field. Although such a preference was certainly demonstrated with their experimental stimuli, in which the faces were cut off below the hairline, this cannot explain face specificity in real life because real heads do not have more elements in the upper half (eyes, nose, ears, and mouth are all at the midpoint or in the lower half).

Another likely sounding possibility is that face specificity could arise through attraction to faces based on infants' prenatal familiarity with their mother's voice. Again, however, the fact that such familiarity is known to exist (Kisilevsky et al., 2003; Sai, 2005) is not sufficient. Any theory based on auditory processing makes the prediction that people born deaf would be prosopagnosic, but this is not the case. The same problematic prediction arises from an explanation (Sinha, Balas, & Ostrovsky, 2007) based on an idea of infant preference for moving stimuli that produce synchronous sound (Sai, 2005).

The best "other factor" proposal of which I am aware is that face specificity could arise from faces being placed close enough to infants to be in focus more often than other stimuli (Kanwisher, pers. comm.). In a recent "baby cam" study that recorded the newborn visual world via a camera attached to the baby's head, Sinha et al. (2007) reported that faces were by far the most common stimuli presented close enough to the baby to be visible, given newborns' inability to perceive high spatial frequencies. The faces-in-focus idea leads to a potentially viable mechanism of inheritance of developmental prosopagnosia (or at least one not refuted by current knowledge): namely, unusually poor or unusually good visual acuity in infancy (or, more facetiously, being born to parents with arms of unusual length). The only evidence possibly against it is the lack of view frequency effects.

Overall, I currently lean toward the experience-expectant-innate-template theory. However, claims of innateness are always difficult to back up in that they require ruling out all alternative experience-only explanations. It remains logically possible that an alternative factor explains all the relevant data.

CONCLUSION

Although the question of the developmental origin of holistic/configural processing remains unresolved, the material reviewed in this chapter has provided a clear answer to the question of whether face processing is different from object processing: Yes, it is. This implies that computational and neural models of face recognition can be restricted to the problem of coding the structural form of faces. It also suggests that a valuable focus of future research would be whether, or how, visual representation of face structure is related evolutionarily to other functions associated more directly with the social importance of faces, including processing of eye gaze direction, facial expression, and face reading in speech and communication.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by Australian Research Council DP0450636. Stefan Horarik assisted with final preparation of the manuscript. Material in the "What Could Explain All These Findings?" and "Consistent Evidence From Neuroimaging and Neuropsychology" sections is based strongly on joint discussions with Nancy Kanwisher and Brad Duchaine and is an expansion of material included in McKone, Kanwisher, and Duchaine (2007). The literature review on objects of expertise in the "Faces, Objects, and the Expertise Hypothesis" section is based strongly on discussions with Rachel Robbins and is an expansion of material previously published in Robbins and McKone (2007). I thank Galit Yovel for comments on a draft. Answers to Figure 10.2: top half is George W. Bush and bottom half is Tony Blair.

NOTES

1. Using higher resolution fMRI, Grill-Spector, Sayres, and Ress (2006) claimed the FFA was not uniformly face selective, reporting that it contained many finer scale voxels that were highly selective for nonface objects, such as sculptures. However, Baker, Hutchinson, and Kanwisher (2007) showed this was due to a mistake in the Grill-Spector et al. statistical analysis: Applying the same analysis to noise-only voxels from outside the brain also apparently revealed highly category-selective voxels.

2. Many people reading Diamond and Carey's paper are confused by the fact that the dog experts did not show better performance with upright dogs than did the novices, but instead showed worse performance inverted. The lack of upright advantage does not mean the experts were not experts: It is presumably attributable to the effect of age differences between the groups on memory. Experts' mean age was 64 years, while novices were young college students. Memory declines across this age range.

REFERENCES

Assal, G., Favre, C., & Anderes, J. P. (1984). Nonrecognition of familiar animals by a farmer: Zooagnosia or prosopagnosia for animals. *Revue Neurologique, 140*, 580–584.

Baker, C., Hutchinson, T., & Kanwisher, N. (2007). Does the fusiform face area contain highly selective subregions for nonfaces? *Nature Neuroscience, 10*, 3–4.

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

Bhatt, R. S., Bertin, E., Hayden, A., & Reed, A. (2005). Face processing in infancy: Developmental changes in the use of different kinds of relational information. *Child Development, 76*, 169–181.

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

Bradshaw, J. L., & Wallace, G. (1971). Models for the processing and identification of faces. *Perception and Psychophysics, 9*, 443–448.

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

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

Carey, S. (1981). The development of face perception. In G. M. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 9–38). London: Academic Press.

Carey, S. (1992). Becoming a face expert. *Philosophical Transcripts of the Royal Society of London: B, 335*(1273), 95–102; discussion 102–103.

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

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology, 16*, 257–269.

Cohen, L. B., & Cashon, C. H. (2001). Do 7-month-old infants process independent features or facial configurations? *Infant & Child Development, 10*, 83–92.

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

de Gelder, B., Bachoud-Levi, A., & Degos, J. (1998). Inversion superiority in visual agnosia may be common to a variety of orientation polarized objects besides faces. *Vision Research, 38*, 2855–2861.

de Haan, M., Humphreys, K., & Johnson, M. H. (2002). Developing a brain specialized for face perception: A converging methods approach. *Developmental Psychobiology, 40*, 200–212.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96*, 57–70.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107–117.
Donnelly, N., & Davidoff, J. (1999). The mental representations of faces and houses: Issues concerning parts and wholes. *Visual Cognition, 6*, 319–343.
Duchaine, B. C., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron, 43*, 469–473.
Duchaine, B. C., & Nakayama, K. (2006). The Cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*, 576–585.
Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia, 41*, 713–720.
Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology, 23*, 714–747.
Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) + (retinotopy) ≈ representation of structure. *Spatial Vision, 13*, 255–264.
Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research, 35*, 2089–2093.
Freire, A., & Lee, K. (2001). Face recognition in 4- to 7-year-olds: Processing of configural, featural, and paraphernalia information. *Journal of Experimental Child Psychology, 80*, 347–371.
Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in encoding configural information: Direct evidence. *Perception, 29*, 159–170.
Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103*, 322–330.
Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a nonmodular account of face processing. *Nature Neuroscience, 6*, 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*, 191–197.
Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*, 1673–1682.
Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception & Performance, 28*, 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*, 2401–2428.
Gilchrist, A., & McKone, E. (2003). Early maturity of face processing in children: Local and relational distinctiveness effects in 7-year-olds. *Visual Cognition, 10*, 769–793.
Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*, 225–255.
Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 1023–1039.
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*, 555–562.
Grill-Spector, K., Sayres, R., & Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neuroscience, 9*, 1177–1185.

Hannigan, S. L., & Reinitz, M. T. (2000). Influences of temporal factors on memory conjunction errors. *Applied Cognitive Psychology, 14*, 309–321.
Hasson, U., Hendler, T., Bashat, D. B., & Malach, R. (2001). Vase or face? A neural correlate of shape selective grouping processes in the human brain. *Journal of Cognitive Neuroscience, 13*, 744–753.
Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23*, 65–74.
Hole, G. J., George, P. A., & Dunsmore, V. (1999). Evidence for holistic processing of faces viewed and photographic negatives. *Perception, 28*, 341–359.
Jolicœur, P. (1985). The time to name disorientated natural objects. *Memory and Cognition, 13*, 289–303.
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*, 4302–4311.
Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B, 361*, 2109–2128.
Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception, 19*, 531–543.
Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science, 14*, 220–224.
Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology, 51A*, 449–473.
Leder, H., & Carbon, C.-C. (2005). When context hinders! Learn–test compatibility in face recognition. *Quarterly Journal of Experimental Psychology, 58A*, 235–250.
Leder, H., & Carbon, C.-C. (2006). Face-specific configural processing of relational information. *British Journal of Psychology, 97*, 19–29.
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*, 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*, 762–768.
Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*, 910–916.
Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review, 111*, 835–863.
Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*, 58–70.
Martini, P., McKone, E., & Nakayama, K. (2006). Orientation tuning of human face processing estimated by contrast matching in transparency displays. *Vision Research, 46*, 2102–2109.
Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255–260.
Mazard, A., Schiltz, C., & Rossion, B. (2006). Recovery from adaptation to facial identity is larger for upright than inverted faces in the human occipitotemporal cortex. *Neuropsychologia, 44*, 912–922.
McKone, E. (2004). Isolating the special component of face recognition: Peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 181–197.

McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 310–327.

McKone, E., & Boyer, B. (2006). Four-year olds are sensitive to featural and second-order relational changes in face distinctiveness. *Journal of Experimental Child Psychology, 94*, 134–162.

McKone, E., Brewer, J. L., MacPherson, S., Rhodes, G., & Hayward, W. G. (2007). Familiar other-race faces show normal holistic processing and are robust to perceptual stress. *Perception, 36*, 224–248.

McKone, E., & Grenfell, T. (1999). Object invariance in naming rotated objects: Individual differences and repetition priming. *Perception and Psychophysics, 61*, 1590–1603.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*, 8–15.

McKone, E., Martini, P., and Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 573–599.

McKone, E., Martini, P., & Nakayama, K. (2003). Isolating holistic processing in faces (and perhaps objects). In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes.* Oxford, England: Oxford University Press.

McKone, E., & Peh, Y. X. (2006). Memory conjunction errors for realistic faces are consistent with configural processing, *Psychonomic Bulletin & Review, 13*, 106–111.

McKone, E., & Robbins, R. (2007). The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition, 103*, 331–336.

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

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition, 14*, 55–73.

Michel, C., Rossion, B., Han, J., Chung, C-S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science, 17*, 608–615.

Mondloch, C. J., Dobson, K. S., Parsons, J., & Maurer, D. (2004). Why 8-year-olds cannot tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features. *Journal of Experimental Child Psychology, 89*, 159–181.

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

Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology, 11*, 219–226.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *Journal of Neuroscience, 26*, 11187–11196.

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

Moscovitch, M., & Moscovitch, D. A. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. *Cognitive Neuropsychology, 17*, 201–219.

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

Op de Beeck, H., Baker, C., DiCarlo, J., & Kanwisher, N. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 26*, 13025–13036.

Paivio, A. (1986). *Mental representation: A dual coding approach.* Oxford, England: Oxford University Press.

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition, 82*, 225–257.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long and A. D. Baddeley (Eds.), *Attention and performance IX.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species specific during the first year of life? *Science, 296*, 1321–1323.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science, 14*, 618–622.

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Developmental Science, 9*, 270–277.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze directions. *Proceedings of the Royal Society of London, B: Biological Sciences, 223*, 293–317.

Peterson, M. A., & Rhodes, G. (2003). Introduction: Analytic and holistic processing—The view through different lenses. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes.* New York: Oxford University Press.

Reed, C., Stone, V. E., Bozova, S., & Tanaka, J. W. (2003). The body-inversion effect. *Psychological Science, 14*(4), 302–308.

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

Reisenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London, B (Suppl.), 271*, S488–S450.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47*, 25–57.

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*, 189–203.

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review, 13*, 499–505.

Robbins, R. (2005). *Face and object processing: What changes with experience?* Unpublished PhD thesis, Australian National University.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition, 88*, 79–107.

Robbins R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioral tasks. *Cognition, 103*, 34–79.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized face-like electrophysiological responses. *Psychological Science, 13*, 250–257.

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development, 14*, 29–50.

Sangrigoli, S., & de Schonen, S. (2004). Effect of visual experience on face processing: A developmental study of inversion and non-native effects. *Developmental Science, 7*, 74–87.

Scapinello, K. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science, 21*, 329–331.

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage, 32*, 1385–1394.

Schwarzer, G. (2002). Processing of facial and nonfacial stimuli in 2- to 5-year-old children. *Infant and Child Development, 11*, 253–269.

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

Sergent, J., & Signoret, J. L. (1992). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex, 2*, 375–388.

Simion, F., Valenza, E., Cassia, V. M., Turati, C., & Umilta, C. (2002). Newborns' preference for up-down asymmetrical configurations. *Developmental Science, 5*, 427–434.

Sinha, P., Balas, B., & Ostrovsky, Y. (2007). *Discovering faces in infancy.* Paper presented at Vision Sciences Society, May 11–16, Florida.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A, 46*, 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., Giles, M., Szechter, L., Lantz, J. A., Stone, A., Franks, L., et al. (1996). Measuring parts and wholes recognition of cell, car, and dog experts: A test of the expertise hypothesis. Unpublished manuscript, Oberlin College, Oberlin, OH.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5*(4), 479–496.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*, B1–B9.

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

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

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9*, 483–484.

Tsao, D. Y., Freiwald, W. A., Totell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311*, 670–674.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.

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

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

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*, 59–85.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*, 1234–1242.

Xu, Y. D., Liu, J., & Kanwisher, N. (2005). The M170 is selective for faces, not for expertise. *Neuropsychologia, 43*, 588–597.

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

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

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience, 18*, 580–593.

Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron, 44*, 889–898.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15*, 2256–2262.

Yovel, G., Paller, K. A., & Levy, J. (2005). *Visual Cognition, 12*, 337–352.

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*, 3802–3811.