Available online at www.sciencedirect.com
ScienceDirect
Cognition 103 (2007) 322–330
COGNITION
www.elsevier.com/locate/COGNIT

Discussion

Should we reject the expertise hypothesis? ☆

Isabel Gauthier *, Cindy Bukach

Center for Integrative and Cognitive Neuroscience, Department of Psychology, Vanderbilt University, Nashville, TN 37203, USA

Received 14 April 2006; revised 1 May 2006; accepted 2 May 2006

Abstract

On the basis of a review of the literature and the results of three experiments with dog experts, Robbins and McKone [Robbins, R. A., & McKone, E. (2006). No face-like processing for objects-of-expertise in three behavioural tasks, *Cognition*] argue that there is little or no evidence supporting an expertise account of the differences in configural processing that are typically observed between faces and non-face objects. In the spirit of a debate that has become overly polarized, we believe that R&M often emphasized relatively unimportant controversial issues at the expense of bigger, more important questions. We also feel that some of R&M's arguments are rooted in methodological confusions that should be clarified because they have implications beyond this specific debate. In this commentary, we first clarify issues surrounding the proper statistical analysis of the composite paradigm, a methodology that is commonly used to assess configural and holistic effects in both face and non-face objects. We then discuss several theoretical issues that we feel are central to the debate regarding accounts of face-specificity. We also briefly review positive evidence for the correlation between measures of behavioral expertise and neural markers of face-selectivity. Unlike R&M, we believe the positive evidence for expertise effects, both behavioral and neural, greatly outweighs evidence stemming from null results and that it clearly motivates the importance of future work on the role of experience in the specialization of visual cortex.

☆ This manuscript was accepted under the editorship of Jacques Mehler.
* Corresponding author. Tel.: +1 615 322 1778; fax: +1 615 322 4706.
E-mail address: isabel.gauthier@vanderbilt.edu (I. Gauthier).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.05.003

Keywords: Faces; Expertise; Holistic; Configural; Composite

1. Introduction

Robbins and McKone (2006, R&M) argue that there is little or no evidence supporting an expertise account of the differences in configural processing that are typically observed between faces and non-face objects. They use two main arguments to support their position. First, they suggest that a large body of peer-reviewed research presenting behavioral expertise effects with non-face objects do not in fact offer any substantive evidence. Second, in a series of their own experiments, R&M fail to find expertise effects in dog experts, despite a convincing replication of several standard effects with faces. In this commentary we will ask whether we should reasonably reject the expertise account of face-selectivity based on evidence and arguments presented by R&M. In the spirit of a debate that has become overly polarized, we believe that R&M often choose to emphasize relatively unimportant controversial issues at the expense of bigger, more important questions. We also feel that some of R&M's arguments are rooted in methodological confusions that should be clarified because they have implications beyond this specific debate. In this commentary, we will refrain from debating each previously published finding at length, and instead encourage the reader to consult the primary peer-reviewed articles. We will also refrain from discussing each of R&M's experiments in detail here other than to make the general observation that a strong claim is made (to abandon a conclusion drawn by several peer-reviewed studies) on the basis of null results. Instead, we propose to shift the face-selectivity debate in the context of broader issues of more general import. First, we will discuss the proper statistical treatment of the composite paradigm, a methodology that is gaining increasing popularity in assessing expertise effects for face and non-face categories. We will then discuss several issues of theoretical relevance that are important in assessing the relationship between expertise effects for different object categories.

2. On the proper statistical treatment of the composite paradigm

One way in which R&M end up concluding that there are no expertise effects in our prior work with non-face objects is to selectively focus on some effects at the expense of others. Of the different tasks they use, R&M argue that the composite paradigm offers the strongest evidence because it provides the purest measure of configural processing for faces. In the composite paradigm, participants are asked to make a judgment (e.g., same–different) on a cued half of the face (top or bottom), while ignoring information in the irrelevant noncued half. The degree to which the irrelevant half influences the judgment is considered an index of configural processing. Regarding this paradigm, R&M make a point to argue that our prior work is based on faulty analyses. In our view, this disagreement stems from a misunderstanding of this paradigm. In this section, we discuss the validity of the claims that can be based on two different versions of the composite task. Beyond the present debate, the argument we develop here is relevant to any study that uses a similar paradigm.

R&M argue that in a composite paradigm, only the results of the same trials are theoretically meaningful. According to them, our approach of combining same and different trials to calculate a sensitivity measure is invalid. In order to understand why this is not the case, it is important to realize that there are two different types of composite paradigms that are currently used in the face processing literature. After Gauthier, Tanaka, and Brown (2006), we call the first type "partial composite design" because it contains half of the conditions in the second paradigm, which we will call "complete composite design". In the partial design (used by R&M), the irrelevant face parts are always different, while the cued parts may be same or different. In addition, the two face parts are shown in an aligned or a misaligned format (or in a variant, in an upright vs. inverted format) (see top half of Fig. 1). A configural effect is typically assessed using the difference between aligned and misaligned trials (or upright vs. inverted), only for those trials in which the cued part is same (Hole, 1994; Hole, George, & Dunsmore, 1999; Le Grand, Mondloch, Maurer, & Brent, 2004).

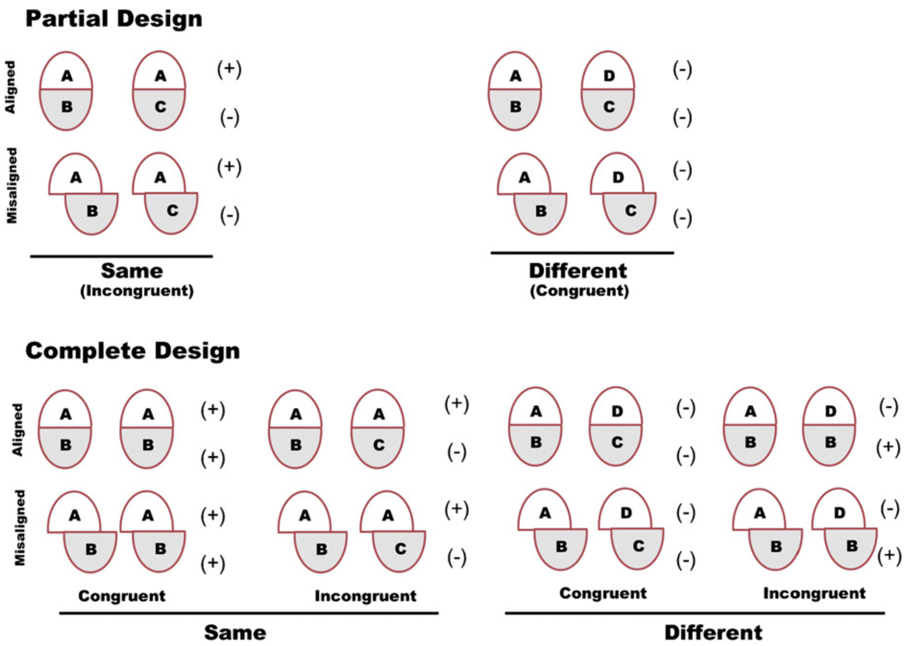

Fig. 1. Schematic diagram depicting partial and complete versions of the composite paradigm. In each face pair, the first face is the study face and the second face is the test face. The cued half (upper) is shown with a white background and the irrelevant face half (lower) is shown with a grey background. In the congruent condition, the study and test face halves (i.e., cued and irrelevant halves) are either both the same ("+") or both different ("–"). In the incongruent condition, the corresponding top or bottom halves of the study and test faces are the same ("+") and the corresponding irrelevant face halves are different ("–").

In contrast, in the complete design, the irrelevant face parts can be same or different *and* the cued parts can also be same or different (Farah, Wilson, Drain, & Tanaka, 1998; Gauthier, Curran, Curby, & Collins, 2003; Gauthier & Tarr, 2002; Wenger & Ingvalson, 2002, 2003) (see bottom half of Fig. 1). In this case, *holistic* processing (a failure of selective attention) is measured using a congruency effect: the congruent trials are those where the cued and irrelevant parts would both lead to identical responses (both same or both different), whereas incongruent trials are those where they are mismatched (one same and the other different). In the complete composite paradigm, the magnitude of the congruency effect is measured by the difference between the congruent and incongruent conditions. This congruency effect is simply a measure of the extent to which the irrelevant part can be ignored. In some studies, but not always, the complete design also includes a manipulation of *configuration*, to test whether the congruency effect is reduced when face parts are misaligned or inverted.

Let's first consider the assertion from R&M that meaningful predictions can only be made for same trials. In the partial design, trials in the "different" condition are always congruent. This means that both the relevant and irrelevant parts are different. In this case indeed, it is difficult to predict whether using information from the irrelevant parts should increase or reduce the perceived similarity of the relevant parts – it depends on whether the two irrelevant parts are more similar to each other than the two relevant parts are. However, in the complete composite paradigm, predictions can be made for the different condition because both congruent and incongruent trials are included. Clearly, when the relevant (cued) parts are different, it will be easier to judge them as such given irrelevant parts that are also different (congruent trials) than irrelevant parts that are identical (incongruent trials). This follows the same logic that is used for the analysis of same trials: when the relevant (cued) parts are same, it will be easier to judge them as same when the irrelevant (noncued) parts are also same (congruent trials), than when the irrelevant (noncued) parts are different (incongruent trials). This demonstrates the benefit of including all possible conditions from the complete composite paradigm and refutes R&M's argument that different trials should not be analyzed due to the inability to make clear predictions.

In fact, there is a more important reason why all conditions are necessary and should be analyzed. It has been demonstrated that observers can have important response biases in the composite paradigm, for instance participants are more likely to respond "same" when the face is upright than inverted (Farah et al., 1998; Wenger & Ingvalson, 2003), or more likely to respond "same" for aligned than misaligned trials (Gauthier et al., 2006). When analyzing only same trials in the partial composite paradigm, it is possible that such a differential response bias for aligned and misaligned conditions is mistaken for a true alignment effect (i.e., a difference in discriminability for aligned vs. misaligned trials). This is demonstrated in the simulated data presented in Table 1. Here, we illustrate the simple case of a subject who shows no congruency effect in any condition: that is, whether the face parts are aligned or misaligned, the absence of a difference between congruent and incongruent trials indicates that the irrelevant part can be successfully ignored. Logically, in this case any valid analysis should conclude that the two face parts are not

Table 1
Simulated data

		same	diff		Hit rate	FA rate	d prime	criterion c	accuracy
ALIGNED									
CONGRUENT TRIALS	"same"	30	5	35	0.6	0.1	2.28	0.51	0.75
	"different"	20	45	65					
INCONGRUENT TRIALS	"same"	**30**	5	35	**0.6**	0.1	2.28	0.51	0.75
	"different"	**20**	45	65					
MISALIGNED									
CONGRUENT TRIALS	"same"	42	15	57	0.84	0.3	2.27	-0.24	0.77
	"different"	8	35	43					
INCONGRUENT TRIALS	"same"	**42**	15	57	**0.84**	0.3	2.27	-0.24	0.77
	"different"	**8**	35	43					

Numbers in bold are the "same"-identity incongruent trials from the partial composite paradigm that R&M advocate focusing on.

processed in a holistic or configural fashion. However, the analysis advocated by R&M (and used by others in the literature – e.g., Goffaux & Rossion, in press; Le Grand et al., 2004), using only the "same"-identity incongruent trials from the partial paradigm, would lead them to falsely conclude to a large configural effect (60% for aligned trials, 84% for misaligned trials). This occurs simply because of a different response bias in the aligned vs. misaligned trials: this simulated subject, just like real Ss tend to do (Gauthier et al., 2006), produces relatively more "different" responses in the aligned than the misaligned condition, *independently* of the identity of the to-be-ignored part. Sensitivity measures, which use both same and different trials, give an estimate of discriminability that is independent of response bias and thus are a better measure of holistic and configural processing. Of course, it is possible that a configural effect measured by same-identity incongruent trials alone might reflect a real configural effect (e.g., a smaller congruency effect in sensitivity for misaligned than aligned trials) and in most studies using upright faces this probably occurs. But it is crucial to understand that discriminability and response bias effects are independent, and that when comparing different groups of subjects such as done by R&M, differences in response biases must be accounted for. Indeed, R&M report biases to respond "same", stronger for experts than novices, and crucially, we do not know how these biases were influenced by alignment. Importantly, the congruency effect in sensitivity, which is a valid measure of whether the irrelevant part can be ignored, is acknowledged by R&M to show expertise effects in our prior work.

In our experience, several dimensions significantly influence response bias in this paradigm, including whether the relevant part is the top or bottom half, the configuration of the parts, and the orientation of the faces. These issues are also relevant to other work in which a different alignment effect is observed between conditions based only on an analysis of the "same-identity" incongruent trials, for instance, a smaller alignment effect for high spatial frequency than low spatial frequency filtered faces (Goffaux & Rossion, in press).

3. On debating the bigger theoretical issues

In our opinion, the so-called "debate" between expertise and modular explanations of special processing for faces has become too polarized for its own good. One symptom of this is the attempt to classify every finding as being in favor of the one or the other account. One example is the idea that evidence for face-specificity, be it in the normal human brain, at the single cell level in animals, or in double-dissociations in brain damaged patients, should be counted against the expertise account. For an intelligent discussion of the origins of face-specificity, it is critical to understand that the vast majority of findings in this field report on face-specific effects but do not in fact speak to the origins of these effects. This large body of work demonstrating that face processing is specialized in the brain is equally consistent with both accounts. Very little work directly addresses where that specialization comes from: the reader interested in this question can turn to expertise studies (see Bukach, Gauthier, & Tarr, 2006, for a review), and, for the alternative hypothesis that there may be an innate bias for face processing, to studies in the newborn (Gauthier, 2006; Johnson, 2005).

Another tedious aspect of this literature, also found in R&M's discussion, is an undue emphasis on questions of the magnitude of expertise effects or on the overlap between the neural substrates of expertise effects with non-face objects and those of face selectivity. Of course, these are valid topics of investigation. However, when it comes to asking what the origins of the specialization for faces may be, and in particular what we can learn from expertise effects on this issue, these questions are not as important as they may seem. Consider the following reasoning: any evidence gathered on the origins of face selectivity in studies of non-face object expertise are indirect. We are merely searching for evidence that something similar to specialization for faces can arise for non-face objects, and are prepared to make the inference that the same mechanism can explain specialization for faces. In this context, the fact that a specific effect (e.g., the inversion effect or BOLD response in the FFA) is smaller for objects of expertise than for faces may simply reflect the greater and earlier experience we have with faces than objects. Similarly, the issue of the exact proportion of single neurons in a cortical area that respond to both faces and objects may be difficult both to measure and to interpret – a more important question has been suggested to be whether the neural substrate for faces and

Table 2
Summary of studies showing correlations between behavioral and neural measures of expertise

Study	Technique	*N*	Behavioral measure	Neural measure	*r*	*P*
Gauthier et al. (2000)	fMRI	6	Car minus bird matching	Cars minus birds in location task in rFFA	0.75	<.05
	fMRI	6	Bird minus car matching	Birds minus cars in location task in rFFA	0.82	<.05
Gauthier and Tarr (2002)	fMRI	5 Ss scanned 5 times	Congruency effect with Greebles during expertise training	Greebles upright minus fixation during matching task in rFFA	0.47	0.02
Gauthier et al. (2003)	ERPs	39	Car minus bird matching	N170 amplitude for cars minus faces in a composite task	0.36	<.05
Xu (2005)	fMRI	5	Car matching	Cars minus objects in location task in rFFA	0.61	0.062
	fMRI	5	Bird matching	Birds minus faces in location task in rFFA	0.74	0.015
Gauthier et al. (2005)	fMRI	7	Car minus bird matching	Spatial frequency filtered (LSF/HSF combined) cars – faces in a composite task in rFFA	0.92	<.05
	fMRI	6	Car minus bird matching	Spatial frequency filtered (LSF/HSF combined) and cars – faces in location and identity tasks in rFFA	0.96	<.005

objects of expertise can function independently of one another (Gauthier et al., 2003; Rossion, Kung, & Tarr, 2004). If expertise with objects can lead to behavioral effects qualitatively similar to those obtained with faces, it is reasonable to postulate that a common mechanism could give rise to both. And let us assume that expertise effects in the brain do not overlap at all with face selectivity in the brain, but are simply found very close to the locus of face selective effects (a conclusion that is very conservative): is it reasonable to infer that the amazing plasticity revealed by these expertise effects with non-face objects has nothing to do with the origins of specialization for faces a few millimeters across in the cortex? It seems that such a conclusion would relegate us to study the brain one neuron at a time, giving up on common principles governing its development.

Finally, although R&M briefly review some evidence for expertise effects in neural studies, their approach is to focus their review and conclusions to behavioral effects. While this divide-and-conquer approach may seem justified given that R&M conducted only behavioral experiments, we would argue that, again in the spirit of focusing on the bigger picture, results that link brain and behavioral evidence are most relevant. R&M do not dispute the existence of effects of expertise in face-selective neural substrates. However, R&M suggest that these effects may be too small to be theoretically significant. Again, it may not be surprising that the magnitude of expertise effects is smaller for objects than for faces with which we have much more experience. However, these effects appear much more impressive when considering the strong correlation obtained between measures of behavioral expertise and neural markers of face-selectivity. For instance, Table 2 shows that in seven independent groups of subjects, behavioral expertise correlates with activity in the FFA in response to non-face objects with a range of $r = .47–.96$. This occurs across a variety of tasks in the scanner, and more importantly, in every case the task and stimuli do not overlap between the fMRI and behavioral measures. It is unlikely that the body of work on this topic is large enough at this point to draw inferences about the true size of this relationship, but these results offer a proof of concept that behavioral expertise relates to cortical activity in (or near, to be conservative) face selective regions. R&M suggest that these effects could be merely attentional (e.g., car experts would respond more to cars because they are more interested in them). Yet, it is unclear why attention to objects should engage the FFA but not other non-face selective visual areas (Gauthier, Curby, Skudlarski, & Epstein, 2005). And it should be noted that in one study (Gauthier et al., 2005), only car novices were recruited and they were scanned 6 months prior to the behavioral measure of expertise, making it unlikely that the strong relationship between performance on matching cars and FFA response ($r = .96$) can be explained by attentional factors to the car stimuli. To be fair, such a tight relationship is not always found and there are a few cases of null findings here too: a salient one is the fact that in Gauthier, Skudlarski, Gore, and Anderson (2000), a correlation was obtained in a location task but not in an identity task.[1] Null effects being what they are, one can only speculate as to why expertise effects were not observed in a particular case (see Bukach et al., 2006, for a review). Unlike R&M, we believe the positive evidence for expertise effects, both behavioral and neural, greatly outweighs evidence stemming from null results and that it clearly motivates the importance of future work on the role of experience in the specialization of visual cortex.

Acknowledgements

This work was supported by a grant from the James S. McDonnell Foundation to the Perceptual Expertise Network. We thank Jim Tanaka for creating (Fig. 1).

References

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends Cognitive Science, 10*, 159–166.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482–498.

[1] Note that the identity task nonetheless revealed a clear expertise effect when comparing car experts to bird experts, and that other studies have found correlations with expertise in identity tasks.

Gauthier, I. (2006). Constraints on the acquisition of specialization for face processing. In Y. Munakata & M. H. Johnson (Eds.), *Attention and performance XXI: processes of change in brain and cognitive development* (pp. 275–284).

Gauthier, I., Curby, K., Skudlarski, P., & Epstein, R. (2005). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive and Affective Behavioral Neuroscience, 5*, 222–234.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*(4), 428–432.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197.

Gauthier, I., Tanaka, J. W., & Brown, D. D. (submitted). Failures of selective attention in the complete face composite paradigm.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 431–446.

Goffaux, V., & Rossion, B. (in press). Faces are "spatial" – Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*.

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

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

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience, 6*, 766–786.

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

Robbins, R. A., & McKone, E. (2006). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*.

Rossion, B., Kung, C. C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 101*(40), 14521–14526.

Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology-Learning Memory and Cognition, 28*(5), 872–892.

Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology-Learning Memory and Cognition, 29*(6), 1106–1118.

Xu, Y. (2005). Revisiting the Role of the Fusiform Face Area in Visual Expertise. *Cerebral Cortex*.