Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 5, 872–892

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.5.872

A Decisional Component of Holistic Encoding

Michael J. Wenger and Erin M. Ingvalson
University of Notre Dame

It has been proposed (see, specifically, M. J. Farah, K. D. Wilson, M. Drain, & J. N. Tanaka, 1998) that human faces are used in cognition as undifferentiated wholes. General recognition theory (GRT; F. G. Ashby & J. T. Townsend, 1986) is used to represent hypotheses regarding the possible sources for the behavioral evidence supporting *holistic representation*. Specifically, it is suggested that holism can be understood in terms of violations of informational independence, informational separability, or decisional separability, as these constructs are defined in GRT. Stimuli were presented upright, inverted, and in an encoding task that emphasized the meaningful nature of the stimuli. Patterns of performance (recognition hit rates) were consistent with prior studies. However, there were only a handful of violations of informational separability. Instead, consistent violations of decisional separability suggested a decisional basis for holistic effects.

One of the most compellingly unitary visual experiences is that of the sight of a human face. A single glimpse of a single face can leave an observer with a myriad of associations, inferences, and judgments. Small wonder, then, that the visual experience of the human face has become the focus of extensive empirical (e.g., Bruce, 1991) and theoretical (e.g., Wenger & Townsend, 2001a) efforts. The present study is concerned with one of the more influential hypotheses regarding facial cognition: the holistic encoding hypothesis.

The most comprehensive articulation of this hypothesis appeared in a recent article by Farah, Wilson, Drain, and Tanaka (1998). In developing their hypothesis, the authors built on and distinguished their ideas from a number of earlier conceptualizations of the encoded face as a unitary psychological entity. Over the past 3 decades, the gestalt character of faces in perception and memory has been attributed to characteristics of both the nature of the underlying representations (the focus of the present study) and the nature of the processes that operate on that information (for a comprehensive discussion, see Farah et al., 1998, pp. 482–485). Constructs pertinent to the nature of the internal representations include second-order or configurational features (e.g., Rhodes, 1988), second-order relational information (e.g., Bartlett & Searcy, 1993; Diamond & Carey, 1986; Searcy & Bartlett, 1996), reliance on low spatial frequencies (e.g., Ginsburg, 1978, 1980), and integrality (as in Garner, 1974) of encoded dimensions (e.g., M. Bradshaw, 1976; Macho & Leder, 1998; Sergent, 1984a, 1984b). Theoretical constructs pertinent to the manner in which these dimensions are processed have typically focused only on the architecture of processing (e.g., parallel vs. serial, as in J. L. Bradshaw & G. Wallace, 1971), although more recent efforts have focused on other critical dimensions of processing (e.g., channel independence, stopping rule, and capacity, as in Wenger & Townsend, 2001b).

The holistic encoding hypothesis falls into the first category of constructs, in that it concerns the nature of the encoded information that supports perceptual judgments and memory performance. As proposed by Farah et al. (1998, p. 484), faces differ from other visual objects because they are based on encoded representations that involve little if any decomposition into component parts, and the processing of faces relies on configural or holistic information to a greater extent than is true for other visual objects. This hypothesis follows from a series of studies (e.g., Tanaka & Farah, 1993; Tanaka & Sengco, 1997) documenting the potent effect that a face context has on the ability to identify and recognize anatomical components of a face (e.g., the nose), with the effect being most pronounced for upright faces.[1]

Of particular importance with respect to the present study is Farah et al.'s (1998) Experiment 1. Observers in that experiment performed same–different judgments on two simultaneously available faces. One anatomical feature was designated as the target feature on each trial, and observers were to make their same–different judgments on that feature. The irrelevant, or nonprobed, features in the face could be either the same as they were in the first stimulus (in which case they were referred to as being *compatible* with the target feature) or different (in which case they were referred to as being *incompatible*). There was a clear advantage for correct "same" responses (i.e., hits) when the irrelevant features were compatible as compared with when they were incompatible, and this advantage was most pronounced when the

Michael J. Wenger and Erin M. Ingvalson, Department of Psychology, University of Notre Dame.

Sincere thanks go to Helena Kadlec for her development and distribution of the software used in the data analyses. We are indebted to Todd Maddox, Jim Tanaka, Angelina Copeland, Alicia Knoedler, Shannon Klatt, and an anonymous reviewer for their generous contributions of constructive comments on earlier versions of this article. Finally, thanks go to Ginny Gerard, Gennifer Gin, Khara Guttierez, David Jones, and Kate McCutcheon for their help in creating and modifying the stimuli.

Correspondence concerning this article should be addressed to Michael J. Wenger, Department of Psychology, 108 Haggar Hall, University of Notre Dame, Notre Dame, Indiana 46556. E-mail: mwenger1@nd.edu

---

[1] One alternative to the holistic encoding hypothesis is the idea that observers use both featural and configural information (e.g., Bartlett & Searcy, 1993; Searcy & Bartlett, 1996). We return to this hypothesis in the General Discussion section.

faces were presented upright rather than inverted (see Farah et al., 1998, Table 1, p. 487).

These results were consistent with the notion that the encoded representations that support the recognition of upright faces involve little if any featural decomposition. Closer inspection of the data, however, reveals that other influences may have been at work. Specifically, because the authors reported accuracy rates in all of the cells of their design, it is possible to analyze performance using the signal detection measures of discriminability (d') and bias (c). When the data are analyzed using these measures, we find both a difference in discriminability as a function of the compatibility of the irrelevant features ($d' = 1.575$ vs. 0.633 for the compatible and incompatible conditions, respectively) and a shift in bias ($c = -0.586$ vs. -0.267 for the compatible and incompatible conditions, respectively). Essentially, when the irrelevant features were incompatible, observers adopted a response criterion that was relatively more conservative compared with when the irrelevant features were compatible. Such a shift suggests that decisional processes may have contributed to effects that have previously been interpreted solely in terms of the characteristics of the encoded representation.

Alternative Representations of the Holistic Encoding Hypothesis

Intrinsic to the holistic encoding hypothesis—both theoretically and in terms of the empirical regularities—is the interaction of information from multiple stimulus dimensions (e.g., information about two anatomical features). For example, the holistic encoding hypothesis suggests that the encoded information about the eyes should be "tightly bound" to the encoded information about the nose and mouth. Thus, the perceptual and mnemonic state of the observer with respect to one feature should be closely related to the perceptual and mnemonic state of the observer with respect to the other features. In this sense, the holistic encoding hypothesis explicitly addresses the simultaneous processing of multiple aspects of a stimulus.

Unfortunately, with the exception of approaches that have relied on methods such as multidimensional scaling (e.g., Takane & Sergent, 1983), the great majority of the evidence collected in support of such interactions has come from tasks in which observers provide reports about only one of the stimulus dimensions (e.g., the nose in the context of compatible or incompatible features, Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). In addition, few studies have explicitly considered the role that decisional processes might play in the obtained effects. Consequently, to provide a comprehensive test of the holistic encoding hypothesis, it would be advantageous to have a way of characterizing the possible dimensional interactions, along with potential decisional influences. In addition, it seems necessary to have an experimental task that would allow observers to provide simultaneous reports about more than one stimulus attribute.

The approach we adopt here uses the multidimensional generalization of signal detection theory known as *general recognition theory* (GRT; e.g., Ashby & Townsend, 1986; Kadlec & Hicks, 1998; Kadlec & Townsend, 1992a, 1992b; Thomas, 1995, 1996). GRT provides a metatheoretical language for representing hypotheses regarding the presence or absence of psychological interactions in the encoded information from multiple stimulus dimensions. The development of GRT was motivated, in part, by the internal inconsistencies in prior approaches to characterizing dimensional relations in internal representations (e.g., integral vs. separable stimulus dimensions; see Maddox, 1992), approaches that have been important in attempts to characterize the facial gestalt (e.g., M. Bradshaw, 1976; Macho & Leder, 1998; Sergent, 1984a, 1984b).

To understand how multidimensional signal detection theory (i.e., GRT) is used in the present study, consider first how a recognition memory task is modeled using classical signal detection theory. After encoding a set of to-be-remembered items (words, pictures, etc.), observers are presented with test items that are either old or new. Presentation of the new items elicits little response from memory, whereas presentation of the old items elicits relatively strong responses (e.g., Atkinson & Juola, 1974; Clark & Gronlund, 1996; Feenan & Snodgrass, 1990; Raaijmakers & Shiffrin, 1981; Ratcliff, 1978; Shiffrin & Steyvers, 1997; Wickelgren & Norman, 1966). Across all items, there will be a distribution of memory responses given new items and a distribution of memory responses given old items. If x is the level of memory response, then we can denote the first distribution (for the new items) as $f_n(x)$ and the second distribution (for the old items) as $f_o(x)$. In signal detection theory terms, these correspond to noise and signal-in-noise distributions, respectively. Given this representation, we assume the observer selects a criterion value of the memory response such that if the obtained level of memory response is above this criterion, an "old" response is generated, and if the obtained level of memory response is at or below this criterion, a "new" response is generated.

Now extend this idea to a situation in which the studied items are composed of more than one critical attribute. Imagine a task in which an observer is asked to encode and remember a facial expression conveyed, for example, by the eyes and nose (such as might be seen in expressions of surprise or disgust). Some time later, the observer is presented with a second view of the face in which either, both, or neither of the two critical features have changed. Presentation of this test stimulus elicits a level of memory strength for each of the dimensions.[2] Assume that the level of memory strength varies from trial to trial within classes of stimuli and within observers (Ashby & Lee, 1993).

Instead of having two different distributions on the level of memory response, we now have four, given two stimulus features, each of which can be old or new. Consequently, our notation needs to become a little more complex. For any one of the four stimulus conditions, let $f_{i,j}(t, m)$ be the multivariate distribution for the level of memory response. Here, the t indicates the level of memory response for the eyes (the top feature) and m indicates the level of memory response for the nose (the middle feature). The subscript i indicates the state of the top feature in the stimulus (o for old, n for new), and the subscript j indicates the state of the middle feature. So, for example, the bivariate distribution of memory response when the eyes of the test face are not different from those of the study face and the nose of the test face is different from the

---

[2] Note that this same conception could apply to the simultaneous same–different task used by Farah et al. (1998, Experiment 1), and the level of evidence could be used to support the same–different judgment in a variety of ways (see detailed discussion in Thomas, 1996).

nose in the study face would be notated as $f_{o,n}(t, m)$. And given that we have two dimensions to consider, we also have two decision criteria: one for the top feature and one for the middle feature.

Representing the situation in this way allows us to consider three constructs from GRT that can be used to characterize the presence or absence of interactions among the internal sources of information for the two stimulus dimensions. The first of these constructs pertains to the information in an individual test stimulus. The memory responses to each of the two stimulus attributes are said to be *informationally independent*[3] if the memory responses to those two attributes are statistically independent, that is,

$$f_{ij}(t, m) = g_i(t)g_j(m),$$

where $g_i(t)$ and $g_j(m)$ are the marginal densities for the two sources of memory information. The holistic encoding hypothesis, in its strongest form, would require a violation of informational independence because if the encoded representation is undifferentiated with respect to the stimulus dimensions, probing the unitary representation with the two cue dimensions should have highly interactive effects. It should be emphasized that informational independence is a construct that applies at the level of each individual stimulus: It would be possible to preserve informational independence when both the eyes and nose are new but to violate informational independence when either or both are old.

A weaker form of the holistic encoding hypothesis can be obtained by considering the second of the theoretical constructs of GRT: *informational separability*. If the two sources of information in the encoded representation are separable, then the effect of a cue at one level of one dimension (e.g., the eyes) does not depend on the level of the cue on the other dimension (e.g., the nose). More specifically,

$$g_{i,o}(t) = g_{i,n}(t),\ i = o, n$$

and

$$g_{o,j}(m) = g_{n,j}(m),\ j = o, n,$$

where $g_{i,j}(t)$ and $g_{i,j}(m)$ are again the marginal densities on the memory response for the top (eyes) and middle (nose) features, respectively. If the marginal densities are instead related by inequalities, then informational separability is violated. This representation of the holistic encoding hypothesis suggests an integrality of information across the possible variations of the stimulus, rather than within a single stimulus.

A violation of either informational independence or informational separability would locate the holism within the encoded (memory) information, either within individual stimuli or across levels of the stimuli. In addition, to be consistent with the notion that the effects are due to encoding, these violations should be present when observers do not need to retain the information in any form of nontransitory memory (i.e., using immediate judgments or a retention interval [RI] of zero, as was the case in Farah et al., 1998, Experiment 1). However, it is also possible, given effects of nonzero RIs on memory for detail (e.g., Anderson & Paulson, 1977; Gernsbacher, 1985; Murphy & Shapiro, 1994; Sachs, 1967; Wenger & Townsend, 2000), that information about the various dimensions of the stimulus could be encoded in an independent manner (e.g., as suggested in Campbell, Schwartzer & Massaro, 2001; Ellison & Massaro, 1997; Massaro, 1998), with violations of independence and/or separability emerging as a function of retention. This would be consistent with the notion of holism in the underlying information but would be inconsistent with the notion that such holism occurs at the time of encoding.

The final theoretical construct of GRT is one whose violation could be capable of producing the compatibility effects described by Farah et al. (1998) but would be inconsistent with the idea of holistic encoding. This construct is *decisional separability*: As it is defined in GRT, decisional separability of the component dimensions is obtained when the decision made about one dimension (i.e., on the basis of the response criterion for that dimension) is unaffected by the level of the other dimension. For example, decisional separability would hold if the criterion for the old–new decision about the eyes did not vary across the old–new status of the nose. Violation of decisional separability by itself would imply that the effects supporting a holistic representation are due primarily to the manner in which observers make decisions about the dimensions and not to the nature of the encoded information.

A point that must be emphasized regarding informational independence, informational separability, and decisional separability (and any violations of them) is that they are properties that do not depend on one level of performance. In this sense, they are analogous to discriminability and bias, as these concepts are measured in classical signal detection theory. Specifically, the values of d' and c depend on the relationship between the hit rates and false-alarm rates, not simply on the level of accuracy. Analogously, inferences about informational independence, informational separability, and decisional separability can be made in the context of shifts in levels of performance, such as might accompany the change from upright to inverted presentation of stimuli.

Overview of the Experiments

There have been two general approaches to applying GRT. The first, widely used in studies of categorization, has been to fit the data of identification and/or confusion matrixes using parameterized models (e.g., Ashby & Alfonso-Reese, 1995; Ashby & Maddox, 1993; Ennis, 1992; Maddox & Ashby, 1993; Maddox & Bogdanov, 2000; Thomas, 1996, 2001). The majority of this work has relied on an assumption of decisional separability that, given the potential for decisional effects in the data from earlier studies (i.e., Farah et al., 1998, Experiment 1), cannot be assumed in the present study (for important related discussions, see Maddox, 2001; Maddox & Bogdanov, 2000). The second approach involves estimating signal detection parameters at two levels of analysis, then using those parameters to guide inferences about the unobservable constructs of informational independence, informational separability, and decisional separability (e.g., Kadlec & Hicks, 1998; Kadlec & Townsend, 1992a, 1992b; Thomas, 2001). This is the approach used in the experiments reported here.

To be able to estimate these signal detection parameters, one must have response data that give information about the observer's

---

[3] Readers familiar with GRT will note that we are taking the liberty of applying the term *informational independence* to the construct of *perceptual independence*. We do this to emphasize that we are concerned with the preservation or violation of independence in the internal representation, or psychological information, specific to the memory task. The construct has the same meaning and application as it would in a perceptual task.

perceptual and mnemonic state with respect to all of the dimensions of concern, not just one. To obtain these data, we designed a set of old–new recognition experiments in which we factorially manipulated the old–new status of two features in a set of objects. We assigned a unique response to each of the four possible stimulus conditions, giving us simultaneous information about the observers' judgments about both dimensions. This experimental design is referred to as a *feature complete factorial design* and the response assignment is referred to as a *complete identification paradigm* (e.g., Ashby & Townsend, 1986; Kadlec & Townsend, 1992b; Kadlec & Hicks, 1998; Townsend, Hu, & Ashby, 1981).

Because we opted for a novel (within this domain) experimental approach, we first needed to substantiate that we could reproduce the effects obtained in earlier work (specifically, Farah et al., 1998, Experiment 1). Second, we needed to determine whether any evidence we might obtain in support of any of the representations of the holistic encoding hypothesis (violations of informational independence, informational separability, and/or decisional separability) were present at encoding or were emergent with the requirement to retain the information in memory. Consequently, in all three experiments, we manipulated the time between presentations of two versions of an image at intervals varying from 0 to 15 s, filling the RI with a backward-counting task (e.g., as in J. Brown, 1958; Peterson & Peterson, 1959). Third, because it has been suggested that holistic representation is something unique to upright human faces, we compared performance with four types of stimuli, all possessing similar types of geometric properties, presented upright and inverted.

Experiment 1 provides a baseline for the studies that follow, establishing that we can obtain effects associated with holistic encoding in upright images using the feature complete factorial design and complete identification response assignment. Experiment 2 involves presenting the images inverted, demonstrating that (among other results) we also can attenuate the effects associated with holistic encoding. Experiment 3 involves a study manipulation intended to have observers treat the images as meaningful wholes at time of encoding, to maximize effects associated with holistic encoding.

Our predictions for the three experiments were as follows. First, we predicted that hit rates (correct "old" responses) and discriminability for any single feature would be higher when the other feature was old rather than new. In addition, we expected that observers would become more conservative in responding to one feature when the other feature was new rather than old. These effects, if obtained, would replicate those obtained by Farah et al. (1998). Second, because the holism that is hypothesized to produce these effects is the result of encoding, we expected to obtain these effects at the shortest RI, with differences remaining across RIs. Third, to the extent that the holistic encoding hypothesis is true, we expected to obtain violations of informational independence or informational separability in the face stimuli starting at the 0-s RI and continuing across all of the RIs. We did not, on the basis of the holistic encoding hypothesis, expect any violations of decisional separability. Fourth, we expected to obtain some violations of informational independence and separability in non–face stimuli but not to the extent that we would obtain them with the face stimuli.

Experiment 1

Method

Participants. A total of 68 participants were recruited from undergraduate psychology courses and were tested in exchange for course credit; 43 contributed data to the analyses reported below. The remaining participants were discarded because false-alarm rates for the top feature of the faces (i.e., the eyes) at the shortest RI exceeded an arbitrary level of 40%.[4] All participants reported normal or corrected-to-normal vision and unencumbered use of their right hand.

Materials. A set of 128 images were obtained (from publicly accessible sites on the World Wide Web) for each of three stimulus categories: human faces, animals, and vehicles. An equal number of images were created for a fourth category, geometric shapes. Equal numbers of male and female faces were selected; each contained an image of an individual face, with hair, but without glasses, hats, or other ancillaries. All faces were cropped at the neck and placed on a black background. Faces were either full-front or 3/4 views. The set of animal images was composed of equal numbers of pictures of domestic dogs and cats, cropped at the neck and placed on a black background. The animal images were either full-front or 3/4 views. The set of vehicle images were either passenger cars or trucks, and were primarily 3/4 (or approximately) views. All humans, scenery, and markings (including markings on the license plates) were removed.

The geometric forms were constructed by placing a white ellipse on a black background. Four elements were placed inside the ellipse, starting with a single black horizontal line near the bottom of the ellipse. The remaining three elements were all identical to each other, and were regular geometric forms (circles, diamonds, squares, etc.) and simple characters (e.g., the % character). These elements were all of the same approximate horizontal and vertical dimensions: two were placed near the top of the surrounding ellipse and the third was placed in the center. The basic form of all the geometric shapes was intended to be face-like, with the upper two elements being placed in the location one would expect for the eyes of a face and the middle element being placed in the location one would expect for the nose.

All of these base images (humans, animals, vehicles, and shapes) were then modified to produce three variations. The first involved increasing the size of the top element (i.e., the eyes in the humans and animals, the headlights in the vehicles, and the two shapes placed in the location of the eyes in the face-like shapes) in vertical direction by between 130.0% and 140.0% of their original size. The actual change was selected so that the variant appeared "natural" to M. J. Wenger (e.g., such as in a change in expression in the faces). Across all of the base images, the mean change involved was 134.7% ($s^2 = 9.7$). The second variant involved increasing the size of the middle element (i.e., the nose in the humans and animals, the blank license plate in the vehicles, and the single shape placed in the location of the nose in the face-like shapes) in the horizontal direction by between 130% and 140%. Again, the actual change was selected so the variant appeared "natural" to M. J. Wenger. Across all of the base images, the mean change was 135.9% ($s^2 = 12.5$). The final variant was created by combining the first and second to produce an image that was modified on both features. All of the images were presented as gray-scale images, at a density of 38 pixels/cm. Each of the images occupied approximately 75% of a 192 × 192 pixel square.

Examples of the stimuli, from all four categories, are presented in Figure 1. Pilot testing on the final set of images indicated that all of the changes supported hit and correct rejection rates of $\geq .80$ at a 0-s RI. To gauge the extent to which the changes in the variants corresponded in

---

[4] The 40% cutoff was set based on pilot work, indicating that with false-alarm rates at this level or higher at the shortest RIs, d's for individual feature recognition at any of the RIs were rarely reliably different from zero.

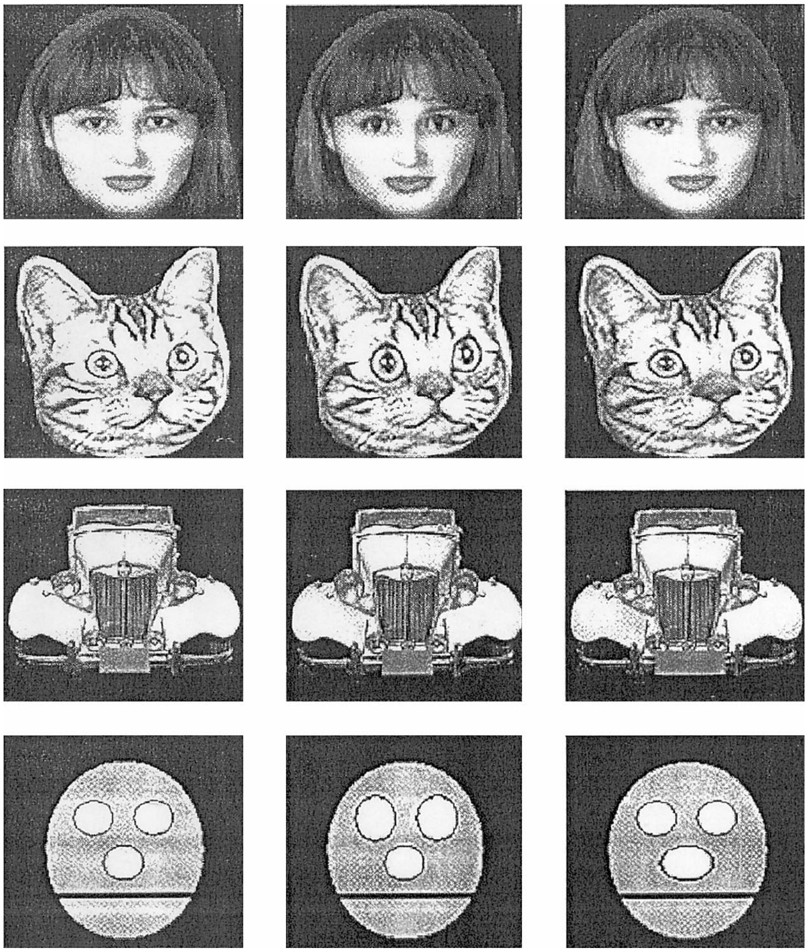

Figure 1. Examples of stimuli from each of the four stimulus classes: humans, animals, vehicles, and shapes. The first column illustrates the base image, the second column illustrates the change in the top feature, and the third column illustrates the change in the middle feature. The image involving changes to both features was composed by combining the images in the second and third columns, and is not shown.

magnitude to those associated with natural changes in expression (in faces), we examined a database of images containing changes in expression. In this analysis (reported in detail in the Appendix), we found that natural changes in expression averaged 167.0% ($s^2 = 64.4$) for increases in the vertical dimension for eyes and 119.0% ($s^2 = 4.9$) for increases in the horizontal dimension for noses.

All stimuli were presented on super video graphics array (SVGA) monitors (33-cm diagonal) controlled by PC-compatible computers. Stimulus onset was synchronized to the vertical refresh of the monitor. All responses were made using the numeric keypad on the right side of the standard PC keyboard and were timed (to ±1 ms) by the PC.

Design. The experiment was conducted as a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 4 (variant: no change, change to the top feature, change to the middle feature, change to both features) complete factorial. All factors were manipulated within observers.

Procedure. Participants were tested in a dimly lit room in groups of between 1 and 5, in sessions lasting 60–90 min. Each trial began with the presentation of a gray fixation cross (approximately the same level of gray as the interior of the faces) for 500 ms at the approximate location of the noses of the human faces. The fixation cross was replaced with the study stimulus, which was present for 3 s. The particular form of the study stimulus on each trial was selected at random from the four possible variants for each stimulus. At the offset of the study stimulus, for all trials involving a nonzero RI, the study stimulus was replaced with a randomly generated three-digit number, presented centered on the screen in a 24-point, bold, white Times Roman font. This number was decremented by a random value between 1 and 10 each second for the duration of the RI. Participants were instructed to count softly out loud along with the decrementing numbers, and the experimenter monitored participants to be sure they complied with this instruction.

At the end of the RI, the test stimulus appeared in the same location as the study stimulus, with the particular form of the test stimulus being determined by the nature of the change being tested on that trial. Participants were instructed to give a simultaneous judgment about the status of the two target features in the test stimulus, as quickly and as accurately as possible, pressing *1* to indicate that they judged both features as old (unchanged), *2* if they judged the top feature as new and the middle feature as old, *3* if they judged the top feature as old and the middle feature as new, and *4* if they judged both features to be new. This response instruction was explained to participants in advance of the testing, and a card summarizing the response assignment was taped to the bottom of each monitor.

Participants were allowed up to 3 s to make a response. If participants failed to respond within 3 s, the trial was logged so as to be discarded prior to analyses. After making their judgment for each test stimulus, participants were presented with the study and the test stimulus alongside each other. This was followed by presentation of the words “The correct response was *x*. Your response was *y*” printed at the same level as the stimuli (with *x* and *y* replaced with the correct and actual responses, respectively). This feedback was present for a total of 1.5 s. Participants were allowed brief breaks every 64 trials.

Results

Although response latency was logged on every trial, we omit analyses of those data to focus on the results pertinent to the questions posed in the introduction. We begin by examining the accuracy data, considering the hit rate, discriminability, and bias data for each of the features separately, to substantiate that we replicated the effects obtained in earlier work on the holistic encoding hypothesis (specifically Farah et al., 1998, Experiment 1). We then present the summary of the GRT analyses aimed at supporting inferences regarding preservation or violations of informational independence, informational separability, and decisional separability.

Accuracy: Hit rates. In Farah et al.’s (1998) Experiment 1, compatibility effects were present in the form of a reduction in the hit rate for the target feature when the irrelevant features changed from compatible to incompatible. Compatibility effects, if obtained in our data, would take the form of a reduction in the probability of a correct “old” response to one feature when the other feature was new rather than old. For completeness, however, we analyzed (separately) the hit rates for the top and the middle feature using a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures analysis of variance (ANOVA). An α level of 0.05 is used here and throughout the article. The results of this analysis are presented in Table 1.

Overall hit rates for each feature across stimulus category, RI, and status of the other feature are presented in Figure 2. In this figure, and many of those that follow, solid lines indicate performance on each of the target features when the other feature was old, and dashed lines indicate performance when the other feature was new. Thus, effects analogous to the compatibility effects reported by Farah et al. (1998) can be seen in the separation between the solid and dashed lines.

Table 1
Results of the Analysis of Variance on Hit Rates for the Top and Middle Features of the Stimuli in Experiment 1

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	17.60	0.058	<.01	26.35	0.065	<.01
Retention interval (R)	3	5.52	0.063	<.01	3.42	0.064	<.05
Other feature (O)	1	6.61	0.065	<.01	3.02	0.058	<.05
C × R	9	3.16	0.059	<.01	2.65	0.056	<.05
C × O	3	7.49	0.077	<.01	7.13	0.063	<.01
R × O	3	1.56	0.065	>.10	1.47	0.067	>.10
C × R × O	9	1.61	0.068	>.10	1.89	0.074	>.10

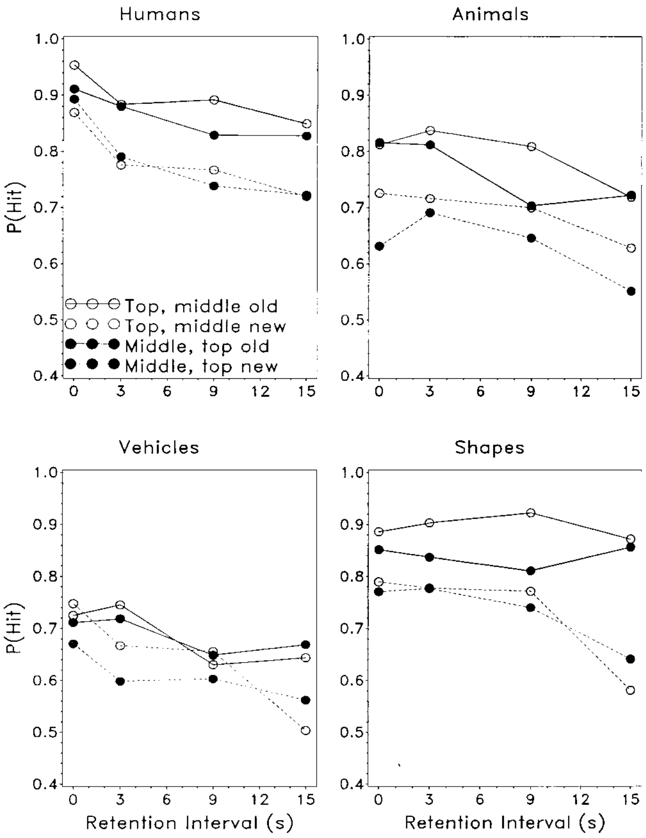

Figure 2. Mean hit rates for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 1 (upright presentation of the stimuli).

As indicated by the data in Table 1, the reduction in hit rates for both the top and middle feature as a function of the other feature being new rather than old was reliable and was not affected by the length of the RI. For the top feature, the effect of changing the middle item from old to new was more pronounced for the humans (0.10 change), the animals (0.06 change), and the shapes (0.08 change) than it was for the vehicles (0.04 change). For the middle feature, the difference in performance as a function of whether the top feature was old or new was greater for the humans (0.10 change) and the shapes (0.07 change) than it was for the animals (0.05 change) or the vehicles (0.03 change).

These results establish that we were able to obtain the effects that have been used as support for the holistic encoding hypothesis when we examined the data for the top and middle features separately. We obtained these results using a very different experimental paradigm from that used in previous studies, and obtaining these results was a critical prerequisite for interpreting the results that follow.

Discriminability. Our examination of the Farah et al. (1998) data revealed a compatibility effect in discriminability (d') as well as in the hit rates. We next examined our data from Experiment 1 to see whether we also obtained compatibility effects in recognition d' for each of the features separately. We examined discriminability for each target feature separately, using a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s)

× 2 (status of the other feature: old, new) repeated measures ANOVA,[5] and the results of this analysis are presented in Table 2. Overall mean discriminability, as a function of item category, status of the other feature, and RI, is presented in Figure 3.

For the top feature, although there was no main effect of the old–new status of the middle item, the difference in discriminability as a function of whether the middle feature was old or new did vary across the RIs. In addition, for the middle feature, although there was no main effect of the old–new status of the top feature, the difference in discriminability as a function of whether the top feature was old or new increased as the RI increased for the humans and the animals but not for the vehicles or the shapes.

The discriminability data thus partially reinforce the findings from the hit-rate data. Discriminability for both the top and middle feature was generally not affected by the status of the other feature. Increasing the RI did serve to produce some reliable differences in discriminability but only for the humans and animals. These effects of RI raise the question of whether the "holism" might be something that emerges with the requirement to retain information in memory.

Bias. Our post hoc examination of the Farah et al. (1998, Experiment 1) data also suggested that there was a shift to a relatively more conservative response bias when the irrelevant feature was incompatible rather than compatible. We thus examined our data to see whether there were any analogous shifts in response criterion. We analyzed the bias data for each of the features separately, using a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures ANOVA, and the results are presented in Table 3. Mean measures of bias for each of the stimulus categories, as a function of status of the other feature and RI, are presented in Figure 4.

For the top feature, observers were reliably more conservative when the middle feature was new rather than old. For the middle feature, observers were also more conservative when the top feature was new rather than old. Thus, the general pattern noted in the data for Farah et al. (1998, Experiment 1), of a shift to a relatively more conservative response bias when the other feature was new rather than old, was obtained in Experiment 1. This was obtained in the context of a novel (to this domain) experimental approach, for both tested features, and for all categories of stimuli.

GRT analyses. Up to this point, we have been examining the data for each of the tested features separately. Although this has allowed us to establish comparability of our data with those of earlier studies, the intent of using the complete identification paradigm was to obtain simultaneous judgments on both stimulus features, so as to assess the data for preservation or violation of informational independence, informational separability, and/or decisional separability. Violations of any of these three properties would be capable of producing the effects associated with holistic encoding, but only violations of informational independence and separability would be consistent with holism in the underlying representation.

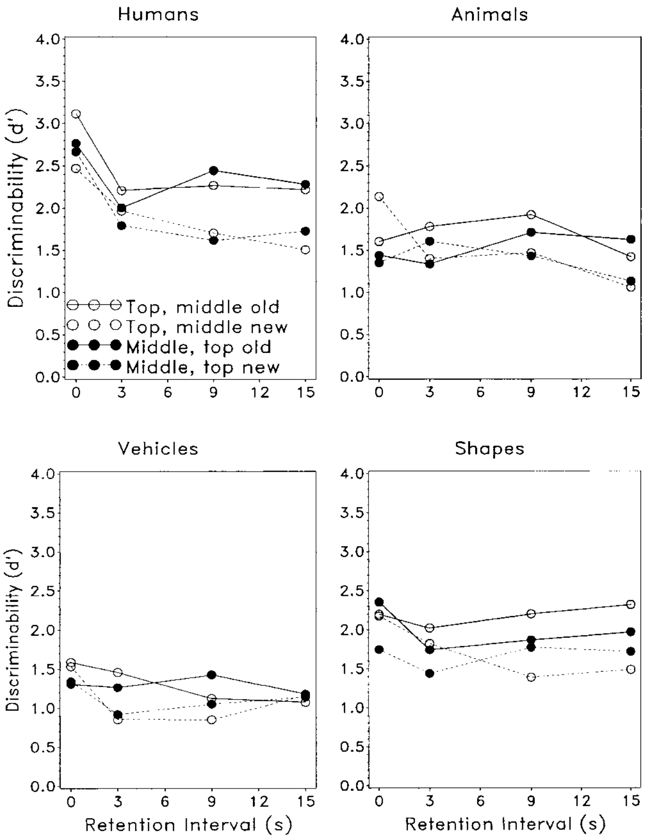

Figure 3. Mean discriminability (in d') for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 1 (upright presentation of the stimuli).

The strategy for making inferences about these unobservable properties is based on testing for a set of observable equivalences in the data. The combined outcomes of these tests point to specific inferences regarding informational independence, informational separability, and decisional separability. The algorithm for using these tests, and the mathematical proofs that support their use, were developed and presented by Kadlec and Townsend (Kadlec

Table 2
Results of the Analysis of Variance on Recognition d' for the Top and Middle Features of the Stimuli in Experiment 1

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	4.77	1.583	<.01	5.02	1.524	<.01
Retention interval (R)	3	23.66	1.770	<.01	34.81	1.696	<.01
Other feature (O)	1	0.28	2.615	>.10	1.81	2.646	>.10
C × R	9	1.03	1.638	>.10	1.18	1.579	>.10
C × O	3	1.31	1.860	>.10	0.68	1.554	>.10
R × O	3	8.34	1.954	<.01	0.03	1.745	>.10
C × R × O	9	0.89	1.720	>.10	2.15	1.793	<.05

---

[5] Although it is not strictly correct to analyze these data using ANOVA, because of the predictable violation of the assumption of homogeneity of variance (see Gourevitch & Galanter, 1967; Macmillan & Creelman, 1991), inferences obtained on the basis of 95% confidence intervals were identical to those supported by the ANOVA, here and in the other two experiments.

Table 3
Results of the Analysis of Variance on Response Bias (c) for the Top and Middle Features of the Stimuli in Experiment 1

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	1.05	0.744	>.10	1.82	0.823	>.10
Retention interval (R)	3	8.12	0.873	<.01	33.57	0.808	<.01
Other feature (O)	1	24.29	1.054	<.01	3.51	1.024	<.05
C × R	9	1.82	0.889	<.10	1.72	0.840	<.10
C × O	3	0.62	0.761	>.10	1.43	0.803	>.10
R × O	3	2.88	0.809	<.10	2.12	0.932	<.10
C × R × O	9	0.73	0.833	>.10	1.25	0.800	>.10

& Townsend, 1992a, 1992b). We summarize the particular tests here and direct readers to Kadlec and Townsend's work for complete explanation of their derivation and use.

The first 3 tests are referred to as *signal detection macro-analyses* and support inferences regarding informational and decisional separability. The first of these tests is referred to as a test of marginal response invariance. For example, marginal response invariance for the top feature (denoted by T) across the levels of the middle feature (denoted by M) holds when

$$P(t_i m_o|T_i M_o) + P(t_i m_n|T_i M_o) = P(t_i m_o|T_i M_n) + P(t_i m_n|T_i M_n), \quad (1)$$

where the subscripts $i, j = o, n$ index the top or middle feature as old or new, respectively. Here and in the definitions that follow, lowercase letters are used to indicate responses and uppercase letters are used to indicate stimuli. Marginal response invariance for the middle feature across the levels of the top feature is defined analogously. If marginal response invariance holds, then there is support for informational and decisional separability. The remaining macro-analyses involve tests for equivalence of marginal d's and βs for each stimulus feature and, along with the results of the tests for marginal response invariance, guide inferences regarding informational and decisional separability. For example, these marginal tests for the top feature would be

$$d'(T, M_o) = d'(T, M_n) \quad (2)$$

and

$$\beta(T, M_o) = \beta(T, M_n), \quad (3)$$

with the tests for the middle feature defined analogously.

The remaining three tests are referred to as *signal detection micro-analyses*. The first of these micro-analyses is a test of sampling independence. If sampling independence. is in force, then the joint probability of reporting a particular conjunction of stimulus states will be equal to the product of each of the marginal probabilities. More specifically,

$$\begin{aligned} P(t_t m_o|T_i M_j) = {} & [P(t_t m_n|T_i M_j) + P(t_t m_o|T_i M_j)] \\ & \times [P(t_n m_o|T_i M_j) + P(t_o m_o|T_i M_j)], \\ & i, j = o, n. \end{aligned} \quad (4)$$

Tests of sampling independence are conducted for all possible stimulus states. The tests for sampling independence are conducted at the level of responses to individual stimuli and thus provide information regarding the preservation or violation of informational independence. If informational independence and decisional separability are preserved, then sampling independence is predicted to hold.

The remaining micro-analyses involve tests for equivalence among measures of discriminability and bias for one feature conditionalized on the level and response given to the other feature. For example, the d' and β for the top feature conditionalized on the responses given when the middle feature is old is

$$d'(T|M = \text{hit}) = d'(T|M = \text{miss})$$

and

$$\beta(T|M = \text{hit}) = \beta(T|M = \text{miss}).$$

All of the other conditionalized measures can defined analogously. These tests are used, along with the results of the tests for sampling independence, to guide inferences regarding informational independence and decisional separability.

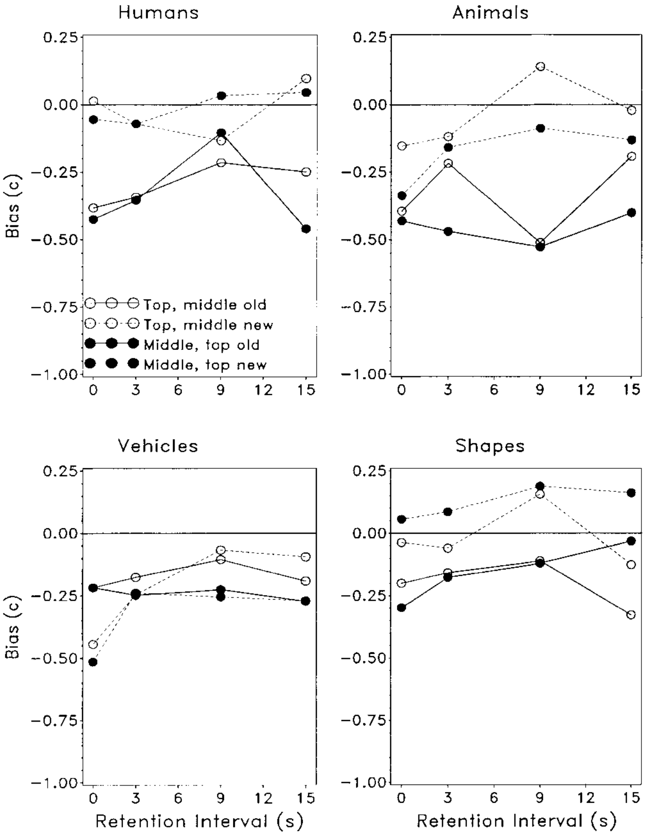

Figure 4. Mean bias (in *c*) for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 1 (upright presentation of the stimuli).

Table 4
Truth Table Relating the Outcomes of the Macro-Analyses for One of the Target Features (e.g., the Top Feature) to Inferences Regarding Informational and Decisional Separability

Evidence			Inferences	
MRI?	Marginal d' equal?	Marginal β equal?	IS	DS
T	T	T	T	T
T	T	F	T	F
T	F	T	F	T
T	F	F	F	F
F	T	T	T	F
F	T	F	T	F
F	F	T	F	?
F	F	F	F	?

Note. This table summarizes details presented in Kadlec and Townsend (1992b; Table 8.1, p. 210). MRI = marginal response invariance; IS = informational separability; DS = decisional separability; T = true; F = false; ? = uncertain inference.

We conducted all of our analyses using software developed specifically for the GRT analyses (MSDA2; Kadlec, 1999).[6] This software implements the macro- and micro-analyses and presents the logical inferences that follow from the results of those analyses. The logic of the inferences is summarized in Tables 4 and 5, which presents the formal relations documented in Kadlec and Townsend's (1992a, 1992b) work.[7] Because of the large number of comparisons and extensive amount of data involved, we restrict our presentation to a graphical summary of the results of the tests, for each of the stimulus types at each of the RIs. A key to the summary is presented in Figure 5. For each of the tests represented in the summary (to follow), a white square indicates that the equivalence held, a gray square indicates that the equivalence did not hold. To the right of each of the summaries (for each stimulus category at each RI), three rectangles indicate the inferences for the top and middle features, with respect to informational independence, informational separability, and decisional separability: A white rectangle indicates that the property held, a gray rectangle indicates the property was violated.

Table 5
Truth Table Relating the Results of the Micro-Analyses to Inferences Regarding Informational Independence and Decisional Separability

Evidence				Inferences		
d'_1 given $H_2 = M_2$	β_1 given $CR_2 = FA_2$	SI in stimuli 1_o2_n 1_o2_o	 1_n2_n 1_n2_o	DS for 1 and 2	II in stimuli 1_o2_n 1_o2_n	 1_n2_n 1_n2_n
T	T	T or F	T or F	T	T	T
T	F	T	T	F	?	?
F	T	T	T	F	?	?
F	F	T	T	F	?	?
T	F	F	F	Either DS or II failed in at least one stimulus, or both. If DS is supported (in macro-analyses), then II failed in at least one stimulus.		
F	T	F	F			
F	F	F	F			

Note. This table summarizes details presented in Kadlec and Townsend (1992b; Table 8.2, p. 211). Results are given for one of the target features (e.g., the top feature, indicated by "1") conditional on the other feature (e.g., the middle feature, indicated by "2"). SI = sampling independence; DS = decisional separability; II = informational independence; H = hit; M = miss; CR = correct rejection; FA = false alarm; subscripts *o* and *n* = the old/new status of the dimension; T = true; F = false; ? = uncertain inference.

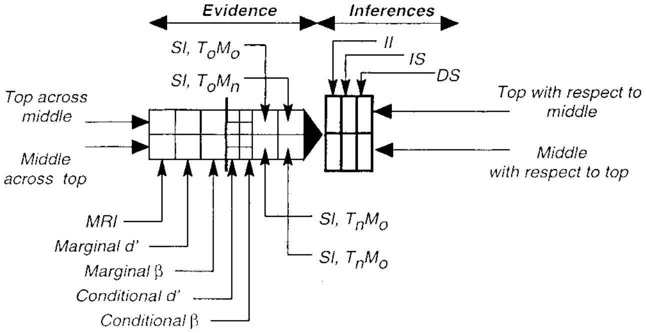

Figure 5. Key to the summaries of the GRT micro- and macro-analyses and supported inferences. If the particular equivalence represented by the cell held in the data, the cell is white; a violation is signaled by a gray cell. Equivalence or nonequivalence of marginal and conditional βs was inferred from tests for equivalence of marginal and conditional values of *c*. MRI = marginal response invariance (Equation 1); Marginal d' = Equation 2; Marginal β = Equation 3; SI = sampling independence (Equation 4); Conditional d' = Equation 5; Conditional β = Equation 5; II = did informational independence hold?; IS = did informational separability hold?; DS = did decisional separability hold?; T_i, $i = o, n$ indexes the top feature as old or new; M_i, $i = o, n$ indexes the middle feature as old or new.

Since Kadlec and Townsend (1992a, 1992b) presented their initial derivations, a number of questions regarding the assumptions supporting the micro-analyses have been raised. It is not clear that violations of these assumptions invalidate any portion of the analyses, and work on these questions is actively being pursued at this writing (specifically, Kadlec, 2000; also, R. D. Thomas, personal communication, August 15, 2000). Consequently, we report the results of the micro-analyses and the inferences they support but suggest caution in interpreting inferences (particularly those based on the micro-analyses alone) as strong evidence.

The summary of the macro- and micro-analyses, and the inferences they supported in Experiment 1, are presented in Figure 6. We obtained consistent violations of decisional separability, for both the top and middle features, for all stimulus categories, at all RIs, excepting the vehicles at the 0-s RI, in which decisional separability was preserved. Informational separability, however, was preserved in all cases, except for the top and middle features

---

[6] This software is available, with documentation, from the University of Victoria's Department of Psychology Web site, http://web.uvic.ca/psyc/.

[7] Of necessity, a great deal of technical information has been omitted in this presentation. Of particular import is the fact that many of the logical relations are directional. in the sense that having one or more theoretical properties in force implies an equivalence in the data, but not the converse. Interested readers are encouraged to consult the tutorial presentations in Kadlec and Townsend (1992b) and Thomas (2001).

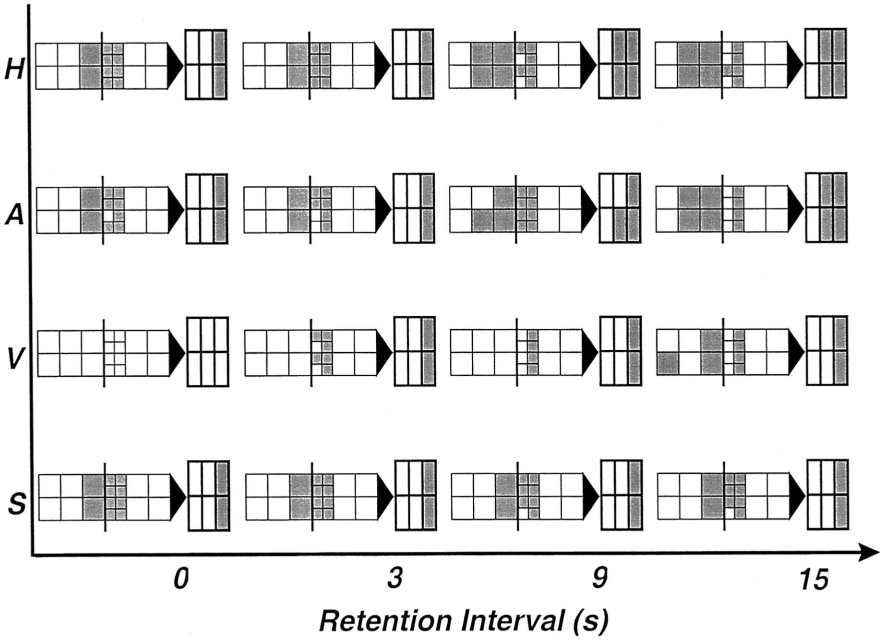

Figure 6. Results of the signal-detection macro- and micro-analyses for the data from Experiment 1 (upright presentation of the stimuli). See Figure 5 for an explanation of the components of the figure. H = humans; A = animals; V = vehicles; S = shapes.

of the humans at the two longest RIs, for the middle feature of the animals at the 9-s RI, and for both features of the animals at the 15-s RI. Thus, we have at best limited support for a weak version of the holistic encoding hypothesis and obtain it only at the longer RIs. With the exception of the vehicles at the shortest RI, we have consistent support for a decisional source at all RIs and for all stimulus categories. There were no violations of informational independence.

Discussion

A critical question for us regarding this first experiment was whether we could obtain the evidence suggestive of holistic encoding in an experimental paradigm requiring simultaneous judgments on two features. Specifically, when our results are examined at a level and in a manner consistent with prior work (Farah et al., 1998), we find the compatibility effects that have been taken as evidence of holistic encoding. Thus, even when observers are required to make simultaneous judgments about two features, rather than a single judgment about one, hit rates for both features are lower when the other feature is new rather than old. In addition, we obtained the shift to a relatively more conservative response bias for one feature when the other feature was new, rather than old, a shift that we noted in Farah et al.'s (1998) data. Consequently, we are reasonably confident that the GRT analyses are providing insight into the effect documented in earlier work.

However, the evidence for holism was not restricted to human faces. The only interactions involving stimulus category were observed in the hit-rate data (see Table 1). On the other measures, we obtained the same general patterns across stimulus categories, even in the context of differences in levels of performance. We obtained these results using stimuli that were all generally left–right symmetric. We should note that some of the non–face stimuli used in previous work (e.g., Farah et al., 1998; Figure 8, p. 493) did not possess this symmetry.

Our analyses of the data using the tools of GRT suggested a consistent violation of decisional separability, with very limited evidence (and this only at the longest RIs) for violations of informational separability. Our data thus point to a distinct decisional basis for those effects that have been interpreted in terms of a unitary encoded representation. Still, we should note that we did obtain some support for violations of informational separability (a weak version of the notion of a holistic representation), and these violations were observed in the stimuli (human faces) that, according to the holistic encoding hypothesis, should be most likely to show such violations. However, these violations of informational separability were obtained at the longest RIs, not the shortest, as would be expected if holistic encoding were occurring. In addition, we obtained similar violations with the animal stimuli, images that should not have supported holistic encoding.

It is possible that given the need to make simultaneous judgments about two features rather than one, the complete identification task may have induced a response strategy that would not allow for holistic encoding. Essentially, observers could have strategically approached the stimuli in such a way as to preserve some type of independence in the encoded representation. At a minimum, this possibility would suggest that performance (at any level of analysis) on either of the features should not show any influence of the old–new status of the other feature. Instead, our hit-rate and bias data show consistent influences. Still, we provide something of a check on this idea in Experiment 3, in which we attempt to directly influence the encoding strategy and encourage observers to treat the stimuli as meaningful wholes.

Experiment 2

Experiment 1 demonstrated that (a) it is possible to obtain the effects associated with holistic encoding in the feature complete factorial design and (b) there may be a distinct decisional (rather than a perceptual, or encoding) basis for these effects. These results were obtained with stimuli that were presented in their normal, upright orientation. This, according to the holistic encoding hypothesis (Farah et al., 1998), is the optimal condition for observing holistic processing. In contrast, inverted presentation should not allow for holistic encoding. Experiment 2 involved presentation of the same stimuli used in Experiment 1, this time presenting those stimuli inverted rather than upright at both study and test.[8] If inversion truly disrupts holistic processing, then we should see an attenuation of the effect of the old–new status of one feature on performance with the other. We should also see preservation of informational separability and decisional separability, and these predictions should hold for all stimulus types.

Method

Participants. Participants were recruited in the same manner as in Experiment 1. Because we are essentially looking for a null effect, we decided to double the sample size relative to Experiment 1. A total of 114 participants were recruited, with 83 contributing data to the final analyses. The remaining 31 participants were discarded using the same criteria applied in Experiment 1.

Materials, design, and procedure. The images used in Experiment 2 were identical to those used in Experiment 1. All that differed was the orientation of the images (inverted rather than upright). Experiment 2 was conducted as a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 4 (variant: no change, change to the top feature, change to the middle feature, change to both features) complete factorial. All factors were manipulated within observers. All details of the procedure were identical to those in Experiment 1, with the exception that all images (both study and test) were presented inverted rather than upright.

Results

Accuracy: Hit rates. The hit rates for the top and middle feature were analyzed separately, using a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures ANOVA, and the results of this analysis are presented in Table 6. The overall results for both features are presented in Figure 7.

As expected, there was very little effect of the old–new status of the middle feature on the hit rates for the top feature. The effect of having the middle feature be new rather than old was more pronounced for the humans (0.07 difference), the vehicles (0.06 difference), and the shapes (0.07 difference) than it was for the animals (0.02 difference). Note that although effect of changing the middle feature from old to new was still significant for some of the inverted stimuli, the magnitude of the effect was attenuated from that observed in Experiment 1. This parallels a similarly attenuated but reliable effect observed with inverted stimuli by Farah et al. (1998, Table 1, p. 487). For the middle feature, the old–new status of the top feature did exert a reliable influence on hit rates for the middle feature. However, the difference between the hit rates when the top feature was old rather than new (0.05) was smaller than the difference observed (0.08) when the stimuli were presented upright (Experiment 1).

Table 6

Results of the Analysis of Variance on Hit Rates for the Top and Middle Features of the Stimuli in Experiment 2

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	13.02	0.070	<.01	17.25	0.077	<.01
Retention interval (R)	3	3.30	0.080	<.05	5.07	0.062	<.01
Other feature (O)	1	2.33	0.113	>.10	4.73	0.066	<.05
C × R	9	3.67	0.071	<.01	3.52	0.070	<.01
C × O	3	12.43	0.089	<.01	2.50	0.097	<.10
R × O	3	0.89	0.090	>.10	0.50	0.066	>.10
C × R × O	9	0.81	0.080	>.10	0.39	0.075	>.10

The results for the hit rates are consistent with our expectation that inversion should attenuate the effects observed in Experiment 1. The results are also consistent with the attenuation observed by Farah et al. (1998) as a function of stimulus inversion. Thus, use of the feature complete factorial design does not seem to have altered the patterns expected on the basis of data from tasks requiring a judgment on only one feature.

Discriminability. The discriminability (d') data were analyzed separately for the top and middle features, using a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures ANOVA. Results of this analysis are presented in Table 7, and the data for both features are presented in Figure 8.

The results for both the top and middle features were consistent with the expectation that inversion would attenuate the influence of effects due to the old–new status of the other feature, relative to when the stimuli were presented upright. For the top and the middle features, no effects involving the old–new status of the middle feature were significant. Thus, for the discriminability data, the small number of effects due to the old–new status of the other feature observed in Experiment 1 disappeared when the stimuli were inverted. These outcomes are consistent with what has been observed in prior work (e.g., Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997), and the attenuation of the influence of the old–new status of one feature on performance with the other when the stimulus is inverted is generally consistent with the evidence used to support the holistic encoding hypothesis.

Bias. The bias measure data (in units of c) for each of the two features were analyzed using separate 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures ANOVA. The results of this analysis are presented in Table 8, and the data for both features are presented in the panels of Figure 9.

For the top feature, observers were relatively more conservative when the middle feature was new than they were when the middle feature was old. For the middle feature, the magnitude of the change in bias as a function of whether the top feature was old or

---

[8] Although Farah et al. did present upright and inverted stimuli within the same experiment, presentation of the stimuli was blocked by orientation (Farah et al., 1998, p. 486).

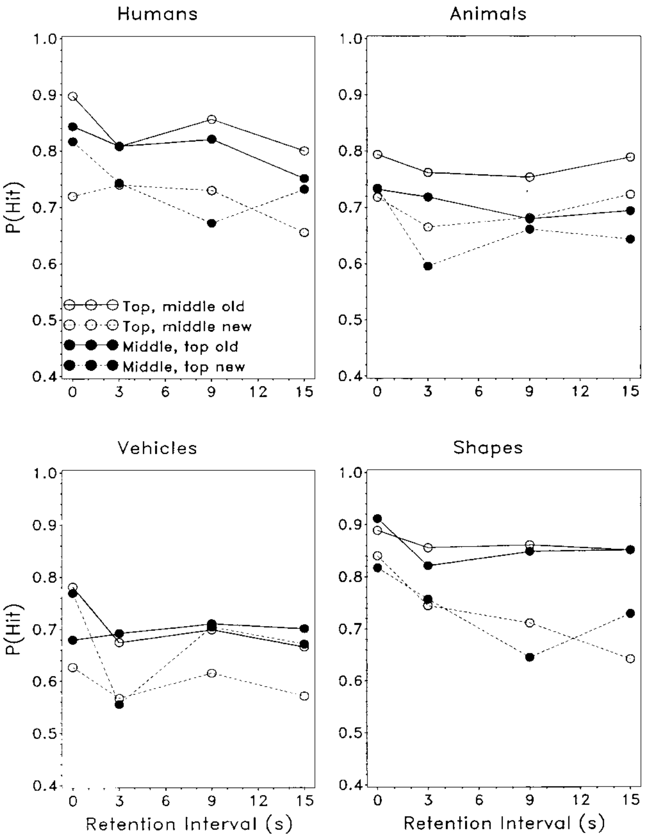

Figure 7. Mean hit rates for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 2 (inverted presentation of the stimuli).

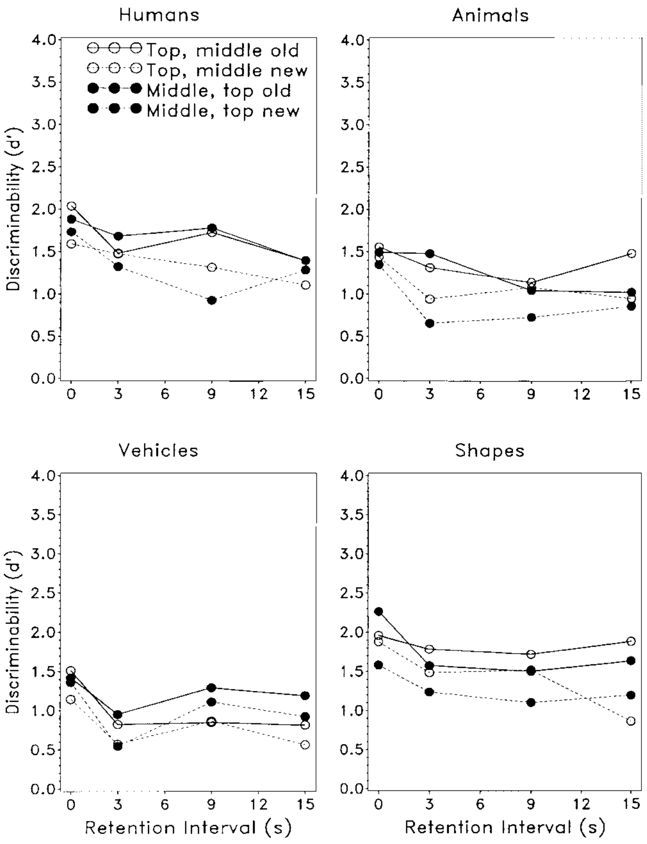

Figure 8. Mean discriminability (in d') for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 2 (inverted presentation of the stimuli).

new was more pronounced for the shapes (0.21) than it was for the humans (0.09), the animals (0.03), or the vehicles (0.06). The bias data from Experiment 2 thus suggest that even when the stimuli are inverted, there is still a shift to a relatively more conservative criterion when the other feature is new rather than old. This shift was also observed when the stimuli were presented upright and in some of the data used to support the holistic encoding hypothesis (specifically Farah et al., 1998, Experiment 1). Essentially, although the level of performance changed in a predictable way when the stimuli were inverted, the effects in the measure of bias were unchanged.

GRT analyses. The signal detection macro- and microanalyses of the data from Experiment 2 are summarized in Figure 10; the key to interpreting this figure is presented in Figure 5. The

Table 7
Results of the Analysis of Variance on Recognition d' for the Top and Middle Features of the Stimuli in Experiment 2

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	7.81	1.712	<.01	10.40	1.996	<.01
Retention interval (R)	3	25.82	1.741	<.01	25.57	1.939	<.01
Other feature (O)	1	0.18	2.029	>.10	2.00	2.618	>.10
C × R	9	1.76	1.814	<.10	2.22	1.580	<.05
C × O	3	1.11	1.794	>.10	1.24	1.890	>.10
R × O	3	1.93	1.782	>.10	1.91	1.891	>.10
C × R × O	9	1.01	1.659	>.10	1.18	1.955	>.10

Table 8
Results of the Analysis of Variance (ANOVA) on Response Bias (c) for the Top and Middle Features of the Stimuli in Experiment 2

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	0.26	0.816	>.10	3.62	0.977	<.05
Retention interval (R)	3	33.57	0.957	<.01	44.68	1.044	<.01
Other feature (O)	1	5.59	1.068	<.05	1.10	1.204	>.10
C × R	9	2.04	0.911	<.05	1.83	0.941	<.10
C × O	3	1.78	0.835	>.10	5.99	0.933	<.01
R × O	3	7.34	0.943	<.01	1.98	0.840	>.10
C × R × O	9	0.54	0.940	>.10	1.14	0.963	>.10

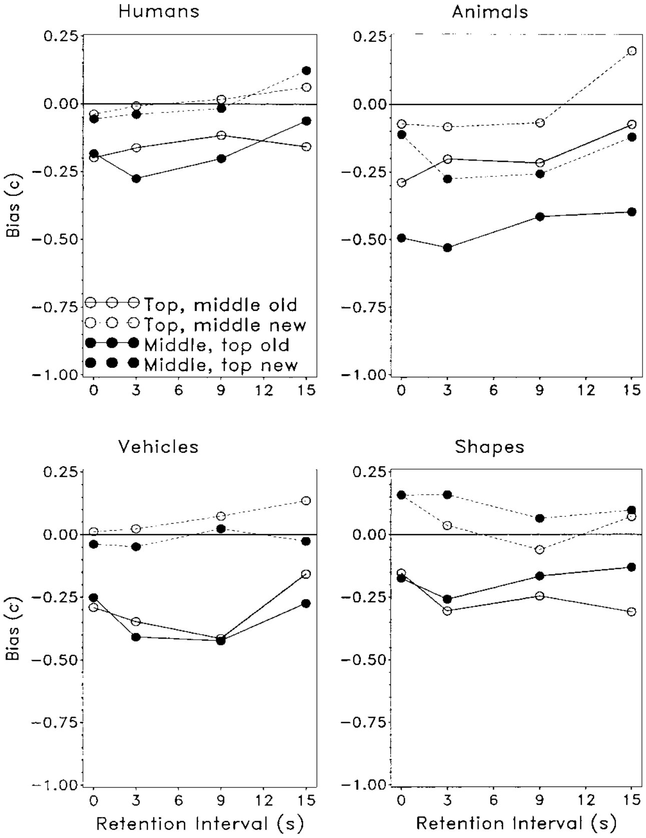

Figure 9. Mean bias (in *c*) for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 2 (inverted presentation of the stimuli).

inferences are rather straightforward and consistent: Decisional separability was violated for all stimulus categories at all of the RIs. There were no violations of informational independence or separability.

Discussion

Experiment 2 was designed to determine whether it would be possible, in the feature complete factorial design, to (a) observe those effects associated with inversion in prior work assessing performance on only one stimulus feature and (b) obtain any changes in the inferences regarding informational and decisional separability in the context of these shifts in performance. With regard to the first of these questions, when performance on each of the target features was analyzed separately, we did find that inversion attenuated the influence of the other feature on both of the target features (relative to performance in Experiment 1). The pattern of attenuated influence is consistent with prior work (specifically Farah et al., 1998, Experiment 1), in which the influence of non–target features was often significant, but reduced in magnitude, when stimuli were presented inverted rather than upright.

In the context of these shifts in levels of performance (obtained for all stimulus categories), observers still evidenced a shift to a relatively more conservative response criterion for individual features when the other feature was new rather than old. This is the same pattern observed in Experiment 1, when stimuli were presented upright. In addition, this pattern is consistent with the presence of a shift to a relatively more conservative response bias (-0.578 to -0.322) in the context of attenuation of the influence of non–target features obtained with inverted stimuli in previous work (see Farah et al., 1998, Experiment 1, Table 1, p. 487).

This apparently pervasive decisional influence was reinforced by the results of the GRT macro-analyses, which showed consistent violations of decisional separability for all stimulus categories at all RIs. Although we found a violation of informational separability at the longest RIs for the human faces in Experiment 1, we found no such violations in Experiment 2. This is generally consistent with the holistic encoding hypothesis, which restricts the holism in the internal representation to upright human faces. In sum, although we were able to attenuate the effects associated with holistic encoding by inversion, we continued to obtain evidence indicating a strong decisional component of performance.

Experiment 3

So far, we have demonstrated that we can, in the feature complete factorial design, reliably obtain the effects that have been used to support the hypothesis of holistic encoding, both in upright and inverted stimuli. In this context, our analyses of performance using the tools of GRT have suggested a strong decisional component and preservation of dimensional separability in the context of findings that have previously been interpreted in terms of a holism in the encoded representation. However, it may be the case that participants in both of the preceding experiments were adopting an encoding strategy that was not optimal for producing a holistic representation. Thus, our goal in Experiment 3 was to induce a strategy that at encoding, emphasized the meaningful nature of the stimuli as wholes, rather than any aspect of surface information (as one would do, e.g., in a levels of processing manipulation, A. S. Brown & Mitchell, 1994; Craik & Lockhart, 1972; Craik & Tulving, 1975; Wenger & Payne, 1997).

Toward that goal, all stimuli in Experiment 3 were presented upright—the condition that should produce a holistic representation, according to the holistic encoding account (e.g. Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). In addition, while participants were studying each of the stimuli, they were required to judge the extent to which a descriptive adjective (one emphasizing an abstract characteristic of the stimulus as a whole) applied to that particular stimulus. If these encoding conditions have their desired effect, we should see an amplification of the effects of manipulating the old–new status of the nontarget feature in separate analyses on each of the target features. Given the persistence of the violations of decisional separability in the inverted stimuli, we would expect to continue to obtain violations of decisional separability. However, to the extent to which the encoding strategy optimizes the possibility for obtaining a holism in the internal representations, we should see more violations of informational separability than we obtained in Experiment 1.

Method

Participants. A total of 95 participants were recruited in the same manner as in the preceding experiments; 80 of these participants contributed data to the final analyses. The rules for discarding participants were identical to those used in Experiments 1 and 2.

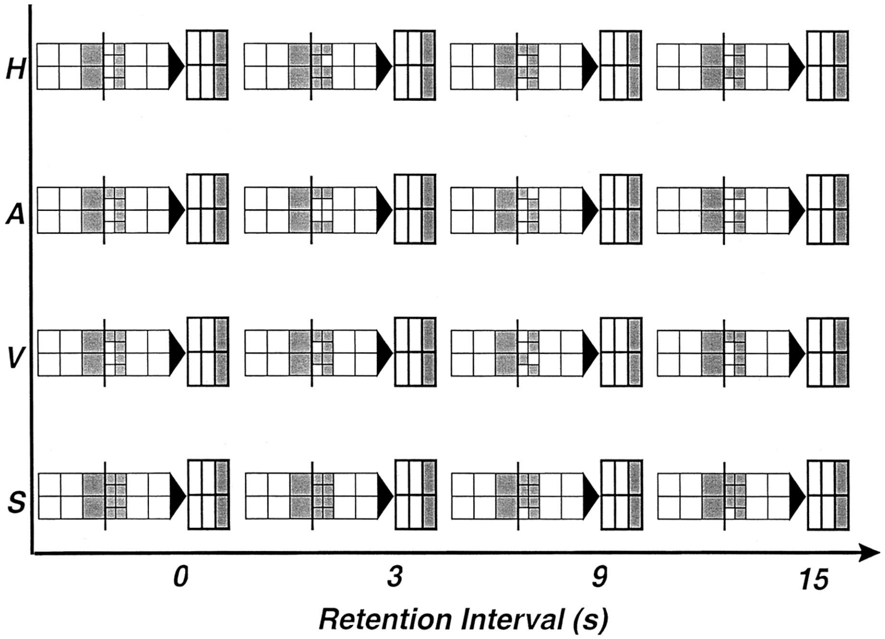

Figure 10. Results of the signal detection macro- and micro-analyses for the data from Experiment 2 (inverted presentation of the stimuli). See Figure 5 for an explanation of the components of the figure. H = humans; A = animals; V = vehicles; S = shapes.

Materials, design, and procedure. The images used in Experiment 3 were identical to those used in Experiments 1 and 2. In addition, eight adjectives (four with a positive evaluative valence and four with a negative valence) were arbitrarily selected for pairing with the images. The adjectives were selected so that (in the judgment of M. J. Wenger) an observer would be unlikely to make a judgment about the property solely on the basis of the top and middle features, alone or independently. These adjectives are listed in Table 9. Experiment 3, like Experiments 1 and 2, was run as a 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 4 (variant: no change, change to the top feature, change to the middle feature, change to both features) complete factorial. All factors were manipulated within observers. All details of procedure were identical to those in Experiment 1, with one exception: When the study item was presented, one of the adjectives (randomly selected with equal likelihood on each trial for each participant) was presented simultaneously above the top left corner of the image in lowercase letters and in the same font used for the backward counting task. Participants were instructed to consider how well the adjective applied to the image, then to give a rating of how well it applied, by using a five-point scale (1 = *does not apply at all* to 5 = *applies very well*). Participants were instructed to use the entire scale and to enter their ratings by using the numeric keypad (these ratings were not recorded).

Results

Accuracy: Hit rates. The hit rates for the top and middle features were analyzed using separate 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures ANOVA. The results of this analysis are presented in Table 10, and the overall results for both features are presented in Figure 11.

Hit rates for the top feature were highest when the middle feature was old rather than new. Note that this is the largest difference in hit rates due to the old–new status of the middle feature observed in all three experiments. In addition, hit rates for the middle feature were highest when the top feature was old rather than new. As was true for the top feature, the difference in hit rates as a function of the old–new status of the top feature was higher in this experiment than it was in either of the preceding experiments. These results suggest that the encoding strategy had its desired effect. Performing the meaningfulness rating at encoding amplified the differences in performance on both of the target features because of the old–new status of the other feature.

Discriminability. Mean levels of discriminability (in units of d') for the top and middle features were analyzed using two separate 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old,

Table 9
Positive and Negative Adjectives Used in the Encoding Task of Experiment 3

Faces, shapes	Animals	Vehicles
	Positive terms	
honest	calm	reliable
friendly	friendly	fun
sincere	playful	agile
calm	affectionate	comfortable
	Negative terms	
angry	angry	stodgy
harsh	scary	difficult
bitter	vicious	irritating
depressed	cold	uncomfortable

Table 10
Results of the Analysis of Variance (ANOVA) on Hit Rates for the Top and Middle Features of the Stimuli in Experiment 3

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	9.10	0.121	<.01	7.08	0.109	<.01
Retention interval (R)	3	0.59	0.095	>.10	2.78	0.090	<.05
Other feature (O)	1	9.14	0.083	<.01	29.12	0.108	<.01
C × R	9	0.98	0.087	>.10	1.00	0.098	>.10
C × O	3	1.69	0.099	>.10	4.37	0.130	<.01
R × O	3	0.18	0.093	>.10	0.89	0.083	>.10
C × R × O	9	1.45	0.091	>.10	0.99	0.087	>.10

Table 11
Results of the Analysis of Variance (ANOVA) on Recognition d′ for the Top and Middle Features of the Stimuli in Experiment 3

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	10.51	2.104	<.01	7.74	1.826	<.01
Retention interval (R)	3	18.09	2.156	<.01	48.74	2.051	<.01
Other feature (O)	1	1.37	1.897	>.10	5.24	1.548	<.05
C × R	9	2.95	1.720	<.01	0.84	1.983	>.10
C × O	3	1.70	1.666	>.10	1.60	1.571	>.10
R × O	3	4.47	2.064	<.01	1.76	1.798	>.10
C × R × O	9	1.26	1.803	>.10	1.56	1.694	>.10

new) repeated measures ANOVA, and the results of this analysis are presented in Table 11. The overall results for both features are presented in Figure 12.

As can be seen in Table 11 and Figure 12, the difference in discriminability for the top feature due to the middle feature being new rather than old was affected by the length of the RI. The old–new status of the top feature had no reliable effect on the discriminability of the middle feature. In spite of the amplification of the differences in performance on each of the target features due to the old–new status of the other feature observed in the hit rates, there were no reliable corresponding differences in overall discriminability. Note that this lack of an effect occurred in the context of overall levels of discriminability that were higher than those obtained in Experiment 1.

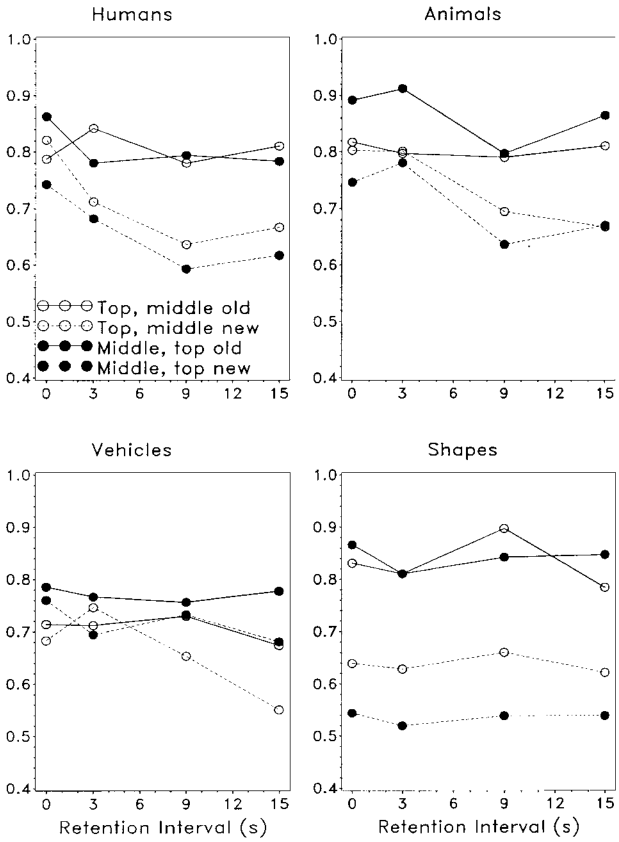

Figure 11. Mean hit rates for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 3 (upright presentation of the stimuli with meaningfulness ratings at encoding).

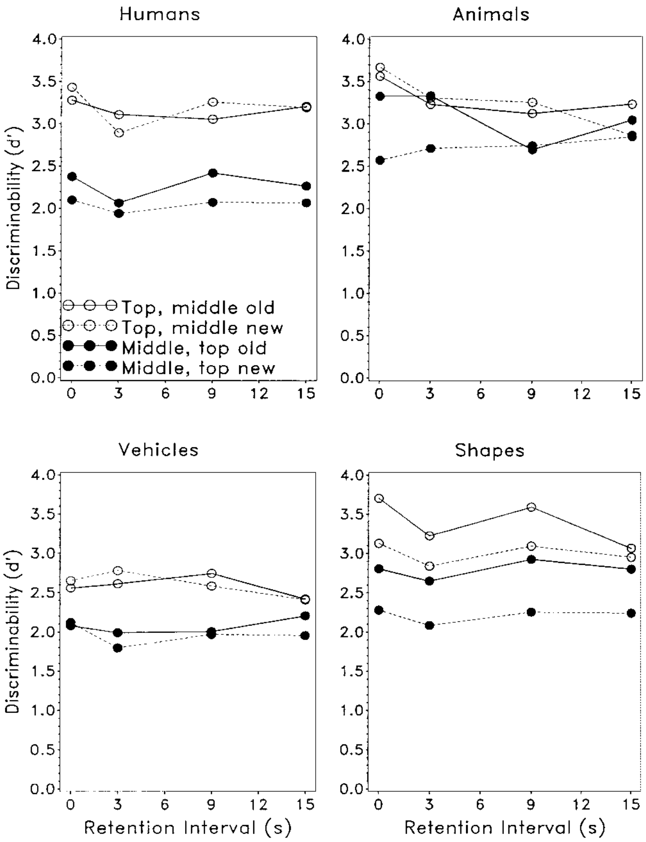

Figure 12. Mean discriminability (in *d*′) for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 3 (upright presentation with meaningfulness ratings at encoding).

Bias. Measures of bias (in units of *c*) for the top and middle features were analyzed using two separate 4 (stimulus category: humans, animals, vehicles, shapes) × 4 (RI: 0, 3, 9, 15 s) × 2 (status of the other feature: old, new) repeated measures ANOVA. Results of this analysis are presented in Table 12, and the data for both features are presented in the panels of Figure 13.

For the top feature, mean bias was relatively more conservative when the middle feature was new (−0.090) rather than old (−0.354). In addition, the impact of the old–new status of the middle feature was dependent on the stimulus category. Differences in bias as a function of the old–new status of the middle feature were larger for the humans (0.170), the animals (0.090), and the shapes (0.122) than they were for the vehicles (0.029). For the middle feature, bias was relatively more conservative when the top feature was new (0.002) rather than old (−0.307). Thus, as was true in both of the preceding experiments, there was a shift to a relatively more conservative response bias on any one of the features when the other feature was new rather than old. This shift was obtained, as it was in the context of the other two experiments, for all of the stimulus categories.

GRT analyses. The signal detection macro-analyses for the data from Experiment 3 are summarized in Figure 14, and the key to interpreting this figure is presented in Figure 5. As was the case in Experiment 2, the results are consistent and straightforward: Decisional separability was violated for all stimulus categories at all RIs, and there were no violations of informational independence or separability, for any of the categories, at any of the RIs.

Discussion

We added the meaningfulness rating to the encoding portion of each trial with the goal of encouraging participants to encode the stimulus as a meaningful whole. Our expectation was that such an encoding task would increase the likelihood of obtaining evidence for a holistically encoded representation (relative to the conditions in Experiments 1 and 2), as evidenced by violations of informational independence or separability.

It was clear that the encoding manipulation did have an impact on performance, as we saw an overall improvement in performance, and observed larger differences in performance (in terms of hit rates) for both of the target features due to the old–new status of the other feature than we did in either of the preceding experiments. In the context of these impacts on the overall level of performance, we continued to obtain a reliable conservative shift in response bias for both of the target features when the other feature changed from old to new. The inferences we made on the basis of the GRT analyses remained consistent with those obtained in the preceding experiments: Performance reflected consistent violations of decisional separability with no violations of informational independence or separability, even at the longest RIs. The implication is that even with a manipulation that directed participants' attention to the whole stimulus at the time of encoding (rather than just the features), a manipulation that had distinct impacts on the differences in levels of performance typically interpreted as evidence for holistic encoding, we failed to obtain any evidence suggesting a violation of separability in the internal representation. Instead, the evidence pointed consistently to a decisional source for the compatibility effects that have been used to support the holistic encoding hypothesis.

Table 12
Results of the Analysis of Variance on Response Bias (c) for the Top and Middle Features of the Stimuli in Experiment 3

		Top feature			Middle feature		
Effect	*df*	*F*	*MSE*	*p*	*F*	*MSE*	*p*
Stimulus category (C)	3	1.49	1.205	>.10	6.96	0.954	<.01
Retention interval (R)	3	8.57	1.037	<.01	50.40	1.102	<.01
Other feature (O)	1	2.84	1.054	<.05	16.45	0.910	<.01
C × R	9	1.77	0.925	<.10	0.86	0.919	>.10
C × O	3	3.42	1.031	<.05	1.92	0.931	>.10
R × O	3	1.88	0.940	>.10	1.27	0.926	>.10
C × R × O	9	1.79	0.865	<.10	1.01	0.874	>.10

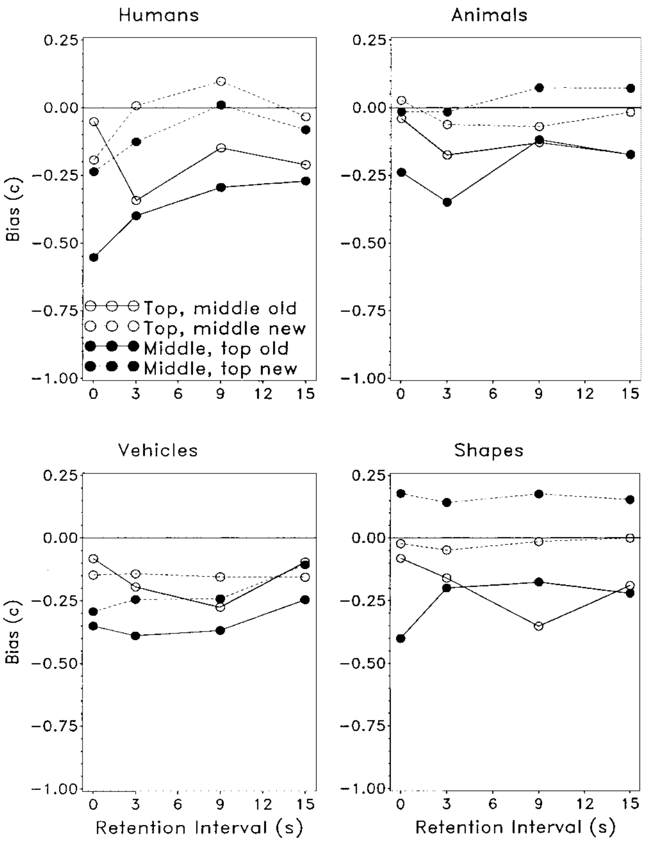

Figure 13. Mean bias (in *c*) for the top and middle features, as a function of stimulus category, retention interval, and the status of the other feature (old or new) in Experiment 3 (upright presentation of the stimuli with meaningfulness ratings at encoding).

General Discussion

The three experiments presented here were designed to probe the hypothesis—compelling both intuitively and on the basis of extant data (Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997)—that upright human faces are perceived, retained,

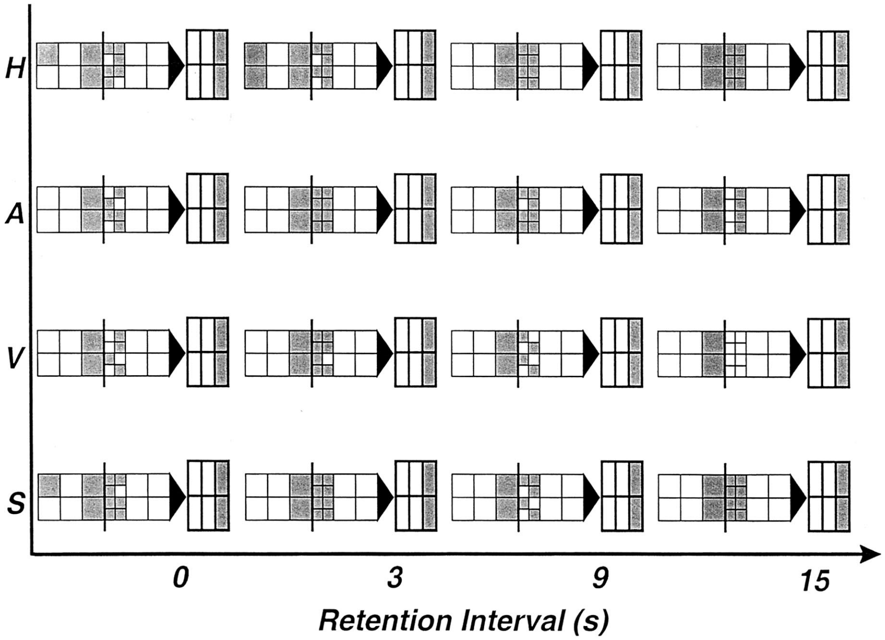

Figure 14. Results of the signal detection macro- and micro-analyses for the data from Experiment 3 (upright presentation of the stimuli with meaningfulness ratings at encoding). See Figure 5 for an explanation of the components of the figure. H = humans; A = animals; V = vehicles; S = shapes.

and used in cognitive tasks as undifferentiated wholes. This holistic encoding hypothesis explicitly addresses itself to the simultaneous use of encoded information from multiple dimensions of a visual stimulus, yet has been supported exclusively by evidence from judgments of a single dimension or anatomical feature. One of the central contributions of the present study is that it provides a direct examination of the simultaneous perception of and memory for more than one dimension of a facial stimulus.

That examination was motivated by considering how a metatheory of cognitive information—GRT (e.g., Ashby & Townsend, 1986; Kadlec & Townsend, 1992a, 1992b; Thomas, 1995, 1996)—might be used to frame alternative representations of the holistic encoding hypothesis. We noted how the three central constructs of GRT—informational independence, informational separability, and decisional separability—could each represent a source for the behavioral regularities that have been used to support the holistic encoding hypothesis. We noted that although violations of any of the three constructs could produce performance consistent with the holistic encoding hypothesis, only violations of informational independence and separability would be consistent with the idea of a holistic internal representation. A violation of decisional separability (particularly in the context of preservation of informational independence and separability) would be capable of producing the behavioral regularities used as evidence for holistic encoding; however, it would be inconsistent with the notion that cognitive actions were based on some type of undifferentiated internal representation.

In order to test the possibilities suggested by GRT, we had to investigate performance by using an experimental paradigm that has not, to our knowledge, been used to examine the holistic encoding hypothesis. That paradigm required participants to give simultaneous old–new responses to two, rather than one, features of the target stimuli. As such, it was critical at each step of the process to ask whether the results we obtained were consistent with previous work. Our data, in all cases, demonstrated that critical consistency: Performance (in terms of hit rates, the measure used to best illustrate holistic encoding; Farah et al., 1998) on each of the target features, when examined separately, was affected by the old–new status of the other feature in the most meaningful stimuli when they were presented upright. That influence was attenuated (but did not disappear) when the stimuli were inverted and was amplified when we encouraged participants to judge the stimuli as meaningful wholes at encoding. In addition, the influence of the old–new status of one feature on the hit rates of the other was present at the shortest RI and did not emerge (or diminish) as the RI increased. Finally, we demonstrated in all three experiments, that the decisional criterion used to guide old–new judgments about each of the target features shifted to something that was relatively more conservative when the other feature was new rather than old. This is a result that was present, although not discussed, in a subset of the data used to support the holistic encoding hypothesis (specifically Farah et al., 1998, Experiment 1). In sum, at the level of the analyses for the individual features, our results are consistent with previous work and with the predictions of the holistic encoding hypothesis.

Given the relative novelty of our approach (at least within the domain of research on face processing), it is worth considering other ways in which our results are consistent with earlier studies. A particular point of note is that inverting the stimuli (in Experiment 2) produced the decreases in performance that have consistently been obtained with inversion (e.g., Bartlett & Searcy, 1993; Farah, Tanaka, & Drain, 1995; Tanaka & Farah, 1991; Valentine, 1991; Yin, 1969). In addition, inversion diminished the impact of the old–new status of one feature on performance (particularly hit rates) with the other feature. This effect was obtained by Farah et al. (1998) as well. Thus, although it is not possible to claim that the observers in our study were doing exactly the same type of perceptual and mnemonic processing as was done by observers in earlier studies, the numerous similarities between our results and those of previous studies give us confidence that we are examining psychological processes that are similar.

However, our data and inferences seem inconsistent with certain physiological results that appear to support the holistic encoding hypothesis (e.g., Kanwisher, McDermott, & Chun, 1997; Kanwisher, Stanley, & Harris, 1999). These studies suggest that the ability to process faces is based on stimulus-specific aspects of cortical organization. By such accounts, one might expect that we should have obtained effects (and thus have supported inferences) with the face stimuli that were qualitatively distinct from those obtained with the other three stimulus classes. However, such reasoning overlooks the fact that the properties of informational independence, informational separability, and decisional separability are independent of the level of performance obtained (as discussed in the introduction). Thus, it is possible to obtain distinct variations in performance across stimulus classes (as we did) while coming to identical conclusions regarding the preservation or violation of informational independence, informational separability, and decisional separability. In that sense, our data are not necessarily inconsistent with studies suggesting specialized cortical processing for faces, because such specialization may lead to distinct levels of performance without predicting distinct patterns of inferences in the GRT constructs. However, there seems to be accumulating evidence for cortical areas that are quite flexible with respect to the type of stimulus they can process and that apply general mechanisms across these different stimuli (e.g., Gauthier, Skudlarski, Gore & Anderson, 2000; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Tanaka & Curran, 2001; Tarr & Gauthier, 2000). We see our results as being more consistent with these latter studies than with the former.

The fact that we were able to obtain effects observed in previous work, in the context of the feature complete factorial design, allowed us to analyze the data with respect to inferences regarding the constructs of GRT. In all three experiments, in the context of the predictable and reliable changes in levels of performance due to inversion and the encoding manipulation, we failed to find any consistent violations of informational separability. Those we did obtain were primarily at the longest of the RIs. This consistent pattern suggests that the dimensions of the internal representation can be treated by observers in an independent manner. Specifically, it means that the psychological information used to make an old–new decision about one of the features in this study did not depend on the old–new status of the other feature.

However, we did observe, for all stimuli and for all RIs, a consistent violation of decisional separability. Essentially, the old–new status of one feature exerted a reliable influence on the manner in which observers set their decisional criteria for judging the other feature to be old or new. Note that all of this occurred in the context of relative levels of performance (i.e., across stimulus classes), and changes in levels of performance (i.e., across stimulus orientation and encoding tasks), that reproduce patterns observed in earlier work involving judgments on only one feature or dimension. Taken together, these inferences suggest that the information in the encoded representation possesses a form of independence (Ashby & Townsend, 1986) and that the effects that have been interpreted as suggesting some type of holism in the internal representation may be due to a regular shift in decisional criteria.

Although it is possible that our results reflect the use of a featural response strategy that may have been induced by the complete identification paradigm, we think that this is not the case. First, the response assignment used in this study assigns a single response to an observer's simultaneous judgment about two elements of the stimulus. These two judgments need to be "pooled" into a single response. A featural response strategy would have been easier and more likely had we asked observers to make separate overt responses for each of the stimulus dimensions. Second, the featural strategy would have been necessary at encoding, not just at the time of test. Had the strategy been applied both at study and test, it is not clear whether the effects of the old–new status of one feature on performance with the other would have been obtained at all. Had the strategy been applied only at test and had the encoded representation been composed of dependent sources of information, violations of informational independence and separability would have been quite likely because the responses would reflect the holism in the originally encoded representation.

The possibility of a feature-based encoding or response strategy raises the question of the extent to which our data speak to one prominent alternative to the holistic encoding hypothesis. This alternative (see, e.g., Bartlett & Searcy, 1993; Cabeza & Kato, 2000; Leder & Bruce, 1998, 2000; Macho & Leder, 1998; Searcy & Bartlett, 1996) posits two sources of information in face processing: one based on the facial features and the other based on the configuration of those features. Tests of this hypothesis have provided evidence suggesting the simultaneous use and importance of both types of information. This challenges the holistic encoding hypothesis in its strongest form, and in that sense, our data are consistent with these studies. However, it should be noted that our results are consistent with results from studies exploring the holistic encoding hypothesis that have indicated the importance of both configural and featural information (e.g., Tanaka & Farah, 1991). Although it is possible that changes in configural information may be the basis for the shifts in response criteria (violations of decisional separability), further modeling and experimentation is necessary to test this possibility.

Our findings thus call into question the strength of the holistic encoding hypothesis (e.g., Farah et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). We found limited or no evidence for holism (as it can be represented using the metatheoretic language of GRT) in the internal representation. However, we should emphasize that we did not call into question any of the empirical regularities that inspired the holistic encoding hypothesis. Our data suggest that these regularities continue to pose important challenges for any theoretical account of visual perception and memory. Our results certainly direct attention to one particular class of mechanism—that being decisional—while de-emphasizing the need for mechanisms that produce some type of gestalt in the internal representation. Indeed, we are currently pursuing work aimed at modeling possible mechanisms for this decisional influence (e.g., as in Ashby, 2000; Smith, 2000; Townsend & Wenger, 1999; Wenger, 1999), as the current set of results document a consistent inference regarding the nature of the underlying information but do not allow any strong inferences regarding processing mechanisms.

The fact that we found almost no evidence for holism in the internal representation (as would be indicated by violations of informational independence and/or separability) also suggests a challenge for additional work. Recent theoretical and empirical work (specifically Kadlec & Hicks, 1998) points to stronger tests for informational separability than those reported here, tests that certainly should be applied before ruling out any violations of perceptual separability. In addition, and as mentioned earlier, fitting parameterized models (as in, e.g., Ashby & Lee, 1991; Maddox & Ashby, 1993; Maddox & Bogdanov, 2000; Thomas, 2001) to the data might provide additional insight into the possibility of any violations of informational separability or independence. In either case, we hope that such research directions, along with the work presented here, demonstrate and emphasize the benefits of extending inquiry and theory in this area by use of pertinent, well-defined theoretical constructs possessing direct links to experimental techniques.

References

Anderson, J. R., & Paulson, P. R. (1977). Representation and retention of verbatim information. *Journal of Verbal Learning and Verbal Behavior, 16,* 439–452.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44,* 310–329.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39,* 216–233.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120,* 150–172.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369–399). Amsterdam: Elsevier.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision-bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review. 93,* 154–179.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1, pp. 243–293). San Francisco: Freeman.

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

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

Bradshaw, M. (1976). *An investigation of stimulus integrality in the perception of schematic faces* (Tech. Rep. No. 82). Baltimore: The Johns Hopkins University, Department of Psychology.

Brown, A. S., & Mitchell, D. B. (1994). A reevaluation of semantic versus non-semantic processing in implicit memory. *Memory & Cognition, 22,* 533–541.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

Bruce, V. (1991). *Face recognition.* Hove, UK: Erlbaum.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science, 11,* 429–433.

Campbell, C. S., Schwartzer, G., & Massaro, D. W. (2001). Face perception: An information-processing perspective. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 285–346). Mahwah, NJ: Erlbaum.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 12,* 599–607.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104,* 268–294.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Ellison, J. W., & Massaro, D. W. (1997). Featural evaluation, integration, and judgment of facial affect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 213–226.

Ennis, D. M. (1992). Modeling similarity and identification when there are momentary fluctuations in psychological magnitudes. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 279–299). Hillsdale, NJ: Erlbaum.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 628–634.

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

Feenan, K., & Snodgrass, J. G. (1990). The effect of context on discrimination and bias in recognition memory for pictures and words. *Memory & Cognition, 18,* 517–527.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

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., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience, 2,* 568–573.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17,* 324–363.

Ginsburg, A. P. (1978). *Visual information processing based on spatial filters constrained by biological data.* (Tech. Rep). Wright–Patterson Air Force Base, Aerospace Medical Research Laboratory, Dayton, OH.

Ginsburg, A. P. (1980). Spatial filtering and visual form perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance. Volume II: Cognitive processes and performance* (pp. 34.1–34.41). New York: Wiley.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika, 32,* 25–33.

Kadlec, H. (1999). MSDA2: Updated version of software for multidimensional signal detection analyses. *Behavior Research Methods, Instruments, & Computers, 31,* 384–385.

Kadlec, H. (2000, August). *What do conditional d's tell us about perceptual independence of dimensions in multidimensional stimuli?* Paper presented at the 33rd Annual Meeting of the Society for Mathematical Psychology, Kingston, Ontario, Canada.

Kadlec, H., & Hicks, C. L. (1998). Invariance of perceptual spaces and perceptual separability of stimulus dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 80–104.

Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology, 36,* 325–374.

Kadlec, H., & Townsend, J. T. (1992b). Signal detection analysis of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–228). Hillsdale, NJ: Erlbaum.

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., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces, not animals. *Neuroreport, 10,* 183–187.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51*(A), 449–473.

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

Macho, S., & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1486–1500.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Erlbaum.

Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception & Psychophysics, 63,* 1183–1200.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision-bound and exemplar models of categorization. *Perception & Psychophysics, 53,* 49–70.

Maddox, W. T., & Bogdanov, S. V. (2000). On the relation between decision rules and perceptual representation in multidimensional perceptual categorization. *Perception & Psychophysics, 62,* 984–997.

Martinez, A. M., & Benavente, R. (1998). *The AR face database* (Tech. Rep. No. 24). West Lafayette, Indiana: Computer Vision Center, Purdue University.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a general principle.* Cambridge, MA: Bradford.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition, 22,* 85–94.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinates of facial appearance. *Perception, 17,* 43–63.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437–442.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904–915.

Sergent, J. (1984a). Configural processing of faces in the left and the right cerebral hemispheres. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 554–572.

Sergent, J. (1984b). An investigation of component and configùral processes underlying face recognition. *British Journal of Psychology, 75,* 221–242.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463.

Stirling (2000). *Psychological image collection at stirling* [Database]. Retrieved from http://pics.psych.stir.ac.uk.

Takane, Y., & Sergent, J. (1983). Multidimensional models for reaction times and same-different judgments. *Psychometrika, 48,* 393–423.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12,* 43–47.

Tanaka, J. W., & Farah, M. J. (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics, 50,* 367–372.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46*(A), 225–245.

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

Tarr, M. J., & Gauthier, I. (2000). Ffa: A flexible fusiform area for subordinate-level visual processing automated by expertise. *Nature Neuroscience, 3,* 764–769.

Thomas, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review, 102,* 192–200.

Thomas, R. D. (1996). Separability and independence of dimensions within the same–different judgment task. *Journal of Mathematical Psychology, 40,* 318–341.

Thomas, R. D. (2001). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contests and challenges* (pp. 193–228). Mahwah, NJ: Erlbaum.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). A test of visual feature sampling independence with orthogonal straight lines. *Psychological Research, 43,* 259–275.

Townsend, J. T., & Wenger, M. J. (1999). Evidence monitoring theory: Foundation and experimental application. 40th Annual Meeting of the Psychonomic Society. Los Angeles, CA: Psychonomic Society Publications.

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

Wenger, M. J. (1999). An initial exploration of the holistic encoding hypothesis: Model and data. 32nd Annual Meeting of the Society for Mathematical Psychology.

Wenger, M. J., & Payne, D. G. (1997). Cue integration across study tasks and direct and indirect retrieval instructions: Implications for the study of retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 102–122.

Wenger, M. J., & Townsend, J. T. (2000). Spatial frequencies in short-term memory for faces: A test of three frequency-dependent hypotheses. *Memory & Cognition, 28,* 125–142.

Wenger, M. J., & Townsend, J. T. (Eds.). (2001a). *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges.* Mahwah, NJ: Erlbaum.

Wenger, M. J., & Townsend, J. T. (2001b). Faces as gestalt stimuli: Process characteristics. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition* (pp. 229–284). Mahwah, NJ: Erlbaum.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3,* 316–347.

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

Appendix

Comparison with Natural Changes in Expression

An important question regarding the magnitude of the changes used in the stimuli, particularly the human faces, is whether they are consistent with the magnitude of changes that one would observe in natural changes of expression. For example, we needed to determine whether the increase in height of the eyes in the faces used in the experiments was consistent with the increase in height of the eyes of a person expressing (for example) surprise. In order to address this question, we examined the magnitude of change in the eyes and nose in a set of facial photographs documenting changes in expression.

Materials

A total of 159 sets of facial photographs used were obtained from publicly accessible databases on the World Wide Web: 90 from the AR Face Database (Martinez & Benavente, 1998) and 69 from the Psychological Image Collection at Stirling (2000). Of these, 141 sets of images were of men and 18 sets were of women. Only frontal views in grayscale were used. The images came in sets of four expressions for each individual; although there was some variance in the type of expressions across individuals, all sets included a neutral expression. This neutral expression was used as the reference face for all changes.

Procedure

All of the measurements were obtained in Adobe Photoshop, with the unit of measurement being one pixel. The height and width of the eyes and nose were found for all of the images in a set. The left eye was measured vertically from eyelid to eyelid and horizontally from corner to corner; the nose was measured vertically from the base to the top of the brow bone and horizontally across the widest part of the nose. Once all the measurements had been taken, those values with the greatest deviation from the measurements found for the reference face were used to calculate the proportion change relative to the reference face for all features in all the faces obtained.

Table A1

Percentage Change to the Eyes (Vertically) and Nose (Horizontally) in Natural Changes of Expression Relative to Images Used in Experiments 1–3

	Natural changes		Stimuli	
Feature	*M* %	*s*	*M* %	*s*
Eyes	161	80	135	10
Nose	116	7	136	13

Results

The proportion change to each of the features across all of the images, in conjunction with the proportion change to each of the features for the stimuli used in Experiments 1–3, are presented in Table A1. We used these data to calculate 95% confidence intervals for the proportion change to each feature for both the stimuli and the natural changes. The overlap of these intervals indicated that the proportion change of the features in the stimuli were not reliably different from what was observed for natural changes in expression.

Received October 11, 2000
Revision received March 11, 2002
Accepted March 11, 2002 ■