State-Trace Analysis of the Face Inversion Effect

Melissa Prince (Melissa.Prince@newcastle.edu.au)
School of Psychology, The University of Newcastle
University Drive, Callaghan, 2308, NSW Australia

Andrew Heathcote (Andrew.Heathcote@newcastle.edu.au)
School of Psychology, The University of Newcastle
University Drive, Callaghan, 2308, NSW Australia

Abstract

The Face Inversion Effect (FIE), the finding that inversion disproportionately affects face recognition, is one of the primary pieces of evidence suggesting that faces are encoded in a qualitatively different way to other stimuli (e.g., along configural as well as featural dimensions). However, when Loftus, Oberg and Dillon (2004) tested the FIE using state-trace analysis (Bamber, 1979), they found evidence for a one dimensional encoding of unfamiliar faces when inversion only occurred during the study phase of a recognition memory test. We report experimental results that replicate Loftus et al.'s findings and rule out several potential problems with their experimental manipulations and state-trace analysis.

Keywords: Recognition Memory; Face Inversion Effect; State-Trace Analysis.

The Face Inversion Effect

There has been a longstanding interest in determining how and why the perception and memory for faces is "special". Humans are expert at recognizing a familiar face after only a glance, when viewed under poor lighting or from a distance and even when seen from a novel viewpoint or in an unfamiliar context. However, our memory is much worse when the faces are unfamiliar (Hancock, Bruce & Burton, 2000) and even more so when they are presented upside-down (Maurer, LeGrand & Mondloch, 2002).

The *Inversion Effect* refers to the robust finding that perception and memory performance for mono-oriented objects (i.e., objects usually viewed in a specific orientation) is strongly disadvantaged by inversion. The *Face Inversion Effect* (FIE) refers to the finding that this inversion effect is disproportionately stronger for faces compared to other mono-oriented stimuli. The FIE is traditionally measured by an interaction comparing the size of the inversion effects for face and mono-oriented control stimuli. It was first reported by Yin (1969), who found an FIE on recognition memory accuracy even when control stimuli (e.g., houses) were matched as closely as possible to faces in terms of complexity, familiarity and difficulty in applying a verbal label. Since Yin's initial demonstration, the FIE in recognition memory has been shown to be a robust phenomenon that has been replicated numerous times and with many procedural variations. Most of these studies have followed Yin's original design where items were studied and tested in the same orientation. However, an FIE has also been observed when all items were studied upright but tested either upright or inverted (Yarmey, 1971) or tested from different viewpoints than study (Valentine & Bruce, 1986).

An apparent exception to these robust findings was recently reported by Loftus, Oberg and Dillon (2004). They found a weak and barely reliable FIE as measured by the standard interaction definition when unfamiliar faces were studied upright or inverted but all tested upright. This design was utilized to examine Valentine's (1988) assertion that to produce an FIE "the orientation of the inspection series does not appear to be critical" (p.474). Indeed Loftus et al. concluded that an FIE would only emerge when inversion is present at the same time as memory retrieval.

Dimensions of the Face Inversion Effect

The FIE has become one of the primary pieces of evidence suggesting that face processing is qualitatively different from the processing of other visual stimuli, that is, that face processing is "special". While the inversion effect is taken to indicate that there is a general factor affecting the processing of all mono-oriented stimuli, the FIE suggests that there is an additional face-specific factor. It has been suggested that the two factors (or dimensions) underlying the FIE might be two types of information, namely featural and configural information. Featural information is common to all mono-oriented stimuli and refers to the isolated features that can be specified without reference to other parts of the object (Rakover, 2002). Configural information, in contrast, is mostly or only available for faces, and at least three different types have been identified (Maurer et al., 2002). The first type, holistic information, captures the overall look of a face (Leder & Bruce, 2000), while the remaining two types refer to the spatial relations between features. First-order information refers to the arrangements of features that define a face (Rhodes, Brake & Atkinson, 1993) and second-order information refers to the distances between internal features (Diamond & Carey, 1986).

While both featural and configural information are affected by inversion, it is usually found that inversion has a greater effect on the availability of configural information (e.g., Leder & Bruce, 2000; Rhodes et al., 1993). Hence it is suggested that upright faces are processed via both featural and configural information, whereas only featural information is available for inverted faces (Carey & Diamond, 1977). Recently, Barton, Keenan and Bass (2001) suggested a more graded view whereby the rate at which both featural and configural information can be extracted is reduced by inversion and that this rate of decrease is stronger for configural information. This graded view also leaves open the possibility that, if given sufficient time, participants may be able to extract some configural information from inverted faces.

Identifying Underlying Dimensions

Evidence for the FIE, and hence the existence of two underlying dimensions for face encoding, is traditionally provided by a dissociation quantified by an interaction test of an accuracy measure. However, such dissociation logic has been shown to be potentially flawed for measures such as accuracy, which can be subject to floor and ceiling effects (e.g., Dunn & Kirsner, 1998; Loftus, 1978). Loftus et al. (2004) proposed state-trace analysis (Bamber, 1979) as a remedy for this flaw. State-trace analysis provides a rigorous method for determining whether a single dimension (i.e., a single latent variable or psychological process) is able to explain the joint effect of two or more experimental factors. Moreover, there is general agreement that state-trace analysis avoids the caveats on traditional dissociation analysis (Newell & Dunn, 2008).

State-trace analysis is most easily explained with reference to a *state-trace plot*, a scatterplot showing the covariation of two factors, a *state factor* and a *dimensional factor*. The state factor defines the axes of the plot. Each point on the plot is defined by a pair of dependent variable values, one for each level of the state factor. Here we use the stimulus type (houses or faces) as the state factor and accuracy as the dependent variable. The dimensional factor, which usually also has two levels, has the potential to differentially affect the dimensionality of the processes determining responses for each state. In our case the dimensional factor is study orientation (upright or inverted) which can potentially cause face processing to change from two to one dimensional.

Dimensionality is determined by whether points on the state-trace plot can be joined by a single monotonic (always increasing or decreasing) function. As at least three points are required to possibly violate monotonicity a third factor, called the trace factor, is usually introduced. This factor must itself have a monotonic effect (i.e., not change dimensionality) and is used to sweep out a "trace" (i.e., set of points) within each level of the dimensional factor. Critically, these traces must overlap for at least one state, otherwise monotonicity will not be violated even if the dimensional factor does change dimensionality. Following Loftus et al. (2004) we used study duration, which can reasonably be assumed to have a monotonic effect on accuracy, as the trace factor.

Using state-trace analysis, Loftus et al. (2004) found evidence for a single dimension (i.e., all points could be joined by a single monotonic line) in accuracy averaged over subjects when they examined memory for unfamiliar faces (experiment 1). However, they found evidence for more than one dimension (i.e., all points could not be joined by a single monotonic line) when the faces were famous (experiment 2; in both cases orientation was manipulated only at study). This evidence led Loftus et al. to conclude that "the FIE emerges when familiar faces are retrieved from memory, but does not emerge when unfamiliar faces are encoded for subsequent recognition" (p.860).

Experiment

Loftus et al. (2004) noted a caveat to their results: they used pictures of famous people but Identikit (drawn) faces as their unfamiliar face stimuli. There is evidence to suggest that configural information is reduced for line drawings compared to photographs of faces (Leder, 1999), and so their use of Identikit faces may have weakened or even removed the FIE in their experiment with unfamiliar faces. Furthermore, differences in stimulus quality from control images may have been an issue as they used photographs as their control stimuli. The current experiment addressed these issues by using photographs in all conditions.

We noted further potential caveats on Loftus et al.'s (2004) results. First, their experiment used the same very brief study durations (17-250ms) for both upright and inverted stimuli. It is possible that brief presentation by itself, causes a failure of configural or holistic encoding as extracting such information is more time consuming than extracting featural information (Palermo & Rhodes, 2002; Valentine, 1988). This caveat is particularly likely for unfamiliar faces, which are more demanding to process than familiar faces, and so they may be more affected by brief presentation. It is also likely to have particularly affected their state-trace results as there was only a very minimal overlap between the traces for upright and inverted conditions (i.e., only the 17ms upright condition overlapped the inverted data). If 17ms were too short to allow extraction of configural information for an upright face the state-trace plot would be monotonic even if a separate configural dimension existed for the longer study durations.

We addressed these caveats in complimentary ways using two between-subjects conditions. Our first condition closely replicated Loftus et al.'s (2004) experiment 1, so will be referred to as the *Test Upright* condition. It differed in that we used longer durations for inverted (i.e., 267-2048ms) than upright (i.e., 33-267ms) stimuli. These values were chosen based on a pilot experiment in order to maximize overlap between upright and inverted traces (i.e., the longer durations for inverted than upright stimuli counteracts the deleterious effect of inversion on accuracy).

We also attempted to use generally longer study durations but were limited in our ability to do so because the *Test Upright* design confounds orientation with the encoding specificity effect. The encoding specificity effect refers to the improvement in memory when study and test conditions match (Tulving & Thomson, 1973). It is a robust and strong effect that has been found to be equal or greater in magnitude than the inversion effect (Rakover & Teucher, 1997). The confounding occurs in the *Test Upright* design because upright items are studied and tested in the same orientation, and so receive a benefit from study-test match as well as from the inversion effect. Inverted items mismatch and so are disadvantaged not only by the inversion effect but also by the encoding specificity effect. As a result, very large differences in study duration are required to compensate for the large deleterious effects on memory for inverted stimuli. Practical limitations did not allow us to use durations much longer than 2s for inverted stimuli, so we had to use shorter study durations for upright stimuli.

Our second condition, which we call the *Test Inverted* condition, attempted to address this issue by testing all items inverted. Pilot testing showed that the study-test match advantage for inverted study stimuli in this design almost exactly counteracted the inversion effect. Hence, we were able to use the same longer set of study durations (267-2048ms) for both orientation conditions.

Method

Participants

Participants (75 in *Test Upright* and 65 in the *Test Inverted* conditions) were recruited from members of the wider community with the only restriction on participation that they had normal or corrected-to-normal vision and were comfortable completing a computer-based task. No demographics were recorded and participants did not receive incentives.

Stimuli

Stimuli were black and white bitmap images (120x105 pixels) displayed at twice their original size. A total of 384 face stimuli were sourced from the FERET database (Phillips, Wechsler, Huang & Rauss, 1998), excluding images with glasses, averted gaze, distinctive facial expressions or natural or photographic blemishes. The faces were divided into homogenous blocks based on race and gender. In total there were 144 African American and 240 Caucasian, with half male and half female. An additional 12 Caucasian male faces were used in a practice phase.

A total of 384 house stimuli (with an additional 12 for practice) were sourced using real estate websites and internet search engines. Houses were excluded if located in New South Wales in order to reduce potential familiarity effects given that participants were drawn from this region. Pilot testing revealed significantly greater accuracy for house than face stimuli. Participant feedback suggested that certain house characteristics made them distinctive within the context of a particular study list. Therefore, house stimuli were presented in homogenous blocks based on their most distinctive feature (e.g., drive-way, fence, etc.).

Apparatus

Testing was completed either at individual computer terminals equipped with 17” LCD monitors or using laptop computers. All stimuli and text were presented on a black background with white font. Prospective and retrospective confidence judgments were made using the computer keyboard with the keys “z”, “x”, “.”, “/” labeled “1”, “2”, “3” and “4” respectively.

Procedure

Testing sessions began with the experimenter reading through the instructions displayed on the participants’ screen. During these instructions it was emphasized that the orientation of a stimulus at study and test was irrelevant to their recognition decision. That is, that they should respond “old” even if the test item was studied in a different orientation. Participants then completed two half-length practice blocks, one using faces and one using houses with order counterbalanced over participants.

The start of a study list was marked by the warning “*Prepare for study. Place your fingers on the keys*” displayed for 2000ms. For each study trial a centrally placed fixation cross was displayed for 1000ms followed by a 300ms blank screen. The target stimulus was then presented for its designated duration following which participants had a maximum of 2500ms to rate their prospective confidence by responding to the question “*How confident are you that you will remember this image later on?*” using a four-point scale (1=“definitely no”, 2=“probably no”, 3=“probably yes” and 4=“definitely yes”). As in Loftus et al. (2004), the purpose of this prospective confidence judgment was to encourage participants to attend to the stimulus and the data from this response was not considered further.

After the study list a 300ms blank screen was followed by the warning “*Prepare for testing. Place your fingers on the keys*”, which appeared for 2000ms. Each test trial was preceded by a 300ms blank screen followed by the test stimulus and retrospective confidence rating scale. The test image was centrally positioned above the question “*How confident are you that you have seen this image earlier?*” and again participants responded using a four-point rating scale where 1=“definitely new”, 2=“probably new”, 3=“probably old”, and 4=“definitely old”. The next trial commenced as soon as the participant responded or if the 5000ms time limit expired. For the entire length of the study and test lists the words “STUDY” and “TEST” were displayed respectively in the top left corner of the screen.

Following practice trials, participants received feedback on the number of times they used each confidence level. The purpose of this feedback was to encourage participants to use the full confidence scale. No feedback regarding accuracy was provided. Participants then commenced the main experiment, which consisted of 32 study-test cycles (16 using face stimuli and 16 using houses). The order of testing face and house stimuli was identical to the practice phase order, such that faces were tested first for half of the participants and houses tested first for the remaining participants. Each study list included 16 images (8 presented upright and 8 inverted), while test lists included 24 images (16 previously studied and 8 new). A 10second break occurred at the end of each cycle and a 5minute break occurred after 16 cycles.

Results

Overall, participants failed to respond on 0.31% of trials. A further 0.64% of test responses were excluded for being faster than 150ms. Accuracy was defined using Loftus et al.'s (2004) "*p*" measure. This measure was obtained by first transforming the 1-4 confidence rating (CR) by *(CR-1)/3*, then averaging to produce for each participant what Loftus et al. refer to as a hit rate (HR) and false alarm rate (FA), where: *p=(HR-FA)/(1-FA).*

We first report a preliminary analysis to ensure the longer durations used in the present study were able to replicate Loftus et al.'s (2004) finding that accuracy was linear as a function of the logarithm of study duration. One-way repeated measures ANOVAs were performed on the effects of the logarithm of study duration for upright and inverted houses and faces in each condition, with polynomial trend analysis. This was followed up by two-way factorial ANOVAs examining the effect of stimulus type (house vs. face) and its interaction with duration.

Linear trends were all highly significant ($p<.001$) and accounted for almost all of the variance in accuracy (see Table 1 and Figure 1). No quadratic or cubic trends approached significance, with the exception of inverted faces (quadratic trend, $p=.04$) and upright faces (cubic trend, $p=.03$) in the *Test Upright* design.

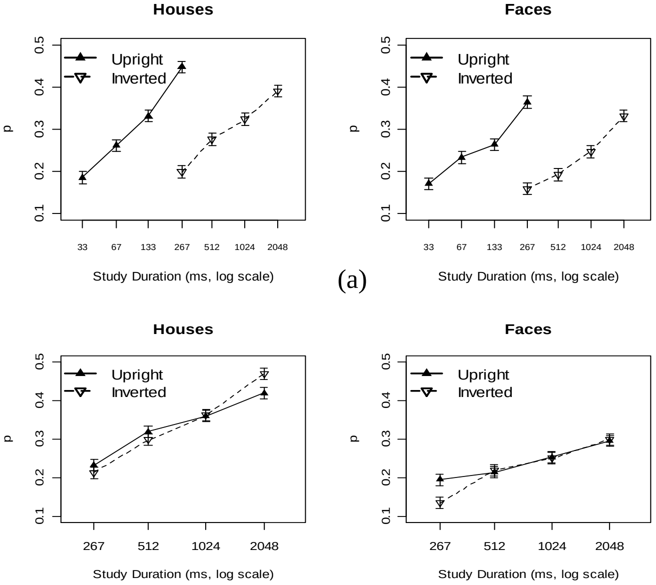
Table 1: Proportion of variance in accuracy accounted for by a linear trend in the logarithm of study duration

	Inverted		Upright	
	Houses	Faces	Houses	Faces
Test Upright	99.2	96.1	98.5	95.4
Test Inverted	99.0	96.1	97.8	97.4

In the *Test Upright* condition accuracy was greater for houses than faces for both inverted (M_H=0.30, M_F=0.23), $F(1,74)=21.66$, $p<.001$, and upright items (M_H=0.31, M_F=0.26), $F(1,74)=12.23$, $p=.001$. Accuracy also increased more quickly with study duration for houses than faces. This effect was reliable for upright, $F(3,222)=3.67$, $p=.01$, but not inverted, $p=.27$. Similarly, in the *Test Inverted* condition, accuracy was higher for houses than faces for both inverted (M_H=0.34, M_F=0.23), $F(1,67)=57.77$, $p<.001$, and upright images (M_H=0.33, M_F=0.24), $F(1,64)=51.21$, $p<.001$. The stronger effect of study duration for houses than faces was reliable for both inverted, $F(3,192)=5.79$, $p=.001$, and upright, $F(2.7,173.8)=3.71$, $p=.02$ (using a Huynh-Feldt correction to degrees of freedom).

We tested for the FIE as traditionally defined by the interaction between orientation and stimulus type. In each condition the corresponding ANOVA used only study durations that were common to upright and inverted conditions. Table 2 also shows, for each duration, estimates of the inversion effect (i.e., difference between upright and inverted) for faces and houses, the corresponding FIE estimates (i.e., inversion effect for faces minus the inversion effect for houses) and the results of associated t-tests.

In the *Test Upright* design, 267ms was the only common duration for upright and inverted items. The FIE was, therefore, tested by a two-way (orientation by stimulus type) ANOVA using only the 267ms data. Accuracy was reliably greater for houses (M=0.32) than faces (M=0.26), $F(1,74)=15.97$, $p<.001$ and for upright (M=0.41) than inverted (M=0.18), $F(1,74)=225.88$, $p<.001$. The FIE interaction test was marginally significant $F(1,74)=3.21$, p=.08, but as shown in Table 2 the effect was in the opposite direction (a greater inversion effect for houses than faces).



(a)

(b)

Figure 1: Accuracy as a function of the logarithm of study duration for (a) *Test Upright* and (b) *Test Inverted* conditions with Loftus and Masson (1994) standard errors.

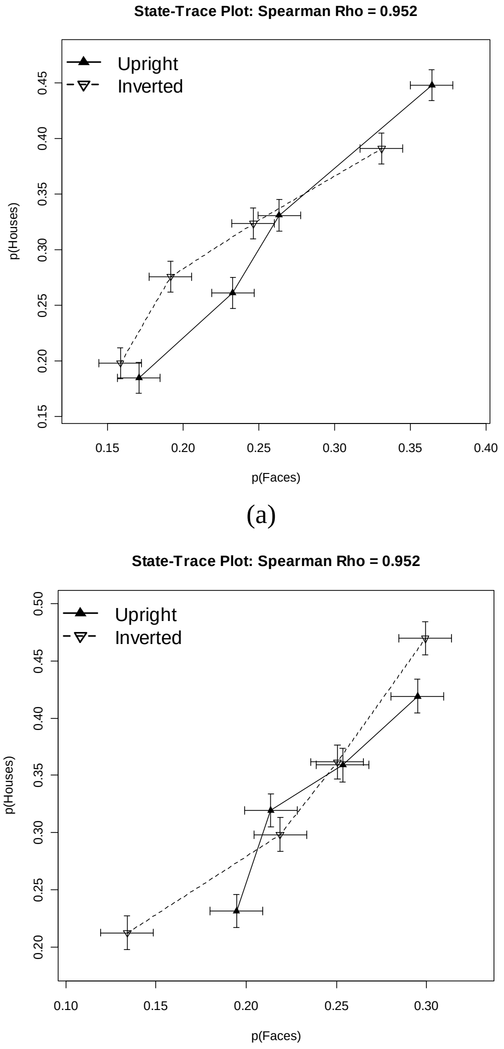
Table 2: Estimates of the inversion effects (IE=Upright-Inverted), face inversion effects (FIE=IE(Faces)-IE(Houses)) and associated t-test results.

	Duration (ms)	IE(Face)	IE(House)	FIE
Test Upright	267	0.206***	0.250***	-0.044
Test Inverted	267	0.061***	0.019	0.041
	512	-0.005	0.021	-0.026
	1024	0.003	-0.003	0.006
	2048	-0.004	-0.051**	0.046

Note: ***=$p<.001$, **=<.01, *=$p<.05$.

The *Test Inverted* condition was fully factorial, so the FIE was examined using a three-way ANOVA that included a duration factor with all four levels. Accuracy was again reliably greater for houses (M=0.33) than faces (M=0.23), $F(1,64)=73.49$, $p<.001$. However, there was no reliable difference in accuracy for upright (M=0.29) and inverted items (M=0.28), $p=.48$. An overall FIE of 0.017 was observed, but the corresponding interaction was not reliable, *p*=.24, and, as shown in Table 2, neither were the FIE estimates at any individual duration.

State-trace plots for each condition are shown in Figure 2. Results for upright study are joined, as are points for inverted study, and these lines are clearly monotonically increasing, consistent with the requirement that the trace factor have a monotonic effect. The plots also show excellent overlap between the two traces in both conditions.



(a)

(b)

Figure 2: State-trace plots for the (a) *Test Upright* and (b) *Test Inverted* conditions with Loftus and Masson (1994) standard errors.

Following Loftus et al. (2004) we examined the monotonicity of the overall plots in two ways. First, we calculated Spearman's ρ, a measure of rank order correlation. Where for ρ=1 perfect monotonicity holds (i.e., the same ordering for points on both axes). Both conditions had the same value of ρ which is close to one because there were only two inversions in the order for each axis. In the *Test Upright* condition these were between the inverted and upright conditions for the two shortest study durations and in the *Test Inverted* condition they were between the middle two durations. The second method involved adding standard errors appropriate for a within-subjects comparison (Loftus & Masson, 1994) to the plots. This aids a visual assessment of whether an inversion is likely to be reliable. For the *Test Inverted* design neither inversion appears reliable as the standard error bars for the inverted points overlap markedly. This is also clearly the case for the lower left pair in the *Test Upright* condition and even for the other pair, where the inversion is more marked, a decrease for inverted houses and an increase for upright houses of less than one standard error would be sufficient to remove the inversion.

Discussion

We replicated Loftus et al.'s (2004) finding of a linear increase in accuracy for study durations up to one quarter of a second and extended this result for durations up to two seconds. The fact that duration effects showed no discontinuity suggests that there is no abrupt change in strategy associated with longer study presentations (i.e., no switch from featural to configural processing). Given evidence that duration and inversion have similar memory effects (Valentine, 1998) the result is consistent with Barton et al.'s (2001) suggestion that inversion does not cause a sudden change in encoding but rather reduces the rate at which featural and configural information are extracted.

Like Loftus et al. (2004) we found little evidence for an FIE using the traditional interaction measure. Although their tests were reliable, the magnitude of their effect was very small (0.042) and was not much different from our results for some durations, which were not statistically reliable. As Loftus et al. (2004) demonstrated in an extensive set of simulations, such inconstancy in the interaction measure of the FIE is to be expected. In contrast, our state-trace results were largely consistent with Loftus et al.'s, although they observed no inversion (ρ=1) whereas we obtained some weak evidence for the occasional inversion. Likely this was due to the much greater overlap between traces in our experiment, which increased the likelihood of chance inversions.

These results are not only consistent with Loftus et al.'s (2004) assertion about the ineffectiveness of study inversion for unfamiliar faces, but also further strengthens this conclusion. First, it shows the prediction is not dependent on the orientation at which items are tested, as our state-trace results were essentially the same when all items were tested upright and when all were tested inverted. Second, the *Test Inverted* condition extended Loftus et al.'s finding to much longer study durations where it is unlikely that insufficient study time was available to perform configural encoding. Finally, through the use of photographic stimuli, the current experiment showed that Loftus et al.'s finding cannot be attributed to their use of Identikit faces.

Despite results consistent with Loftus et al.'s (2004) conclusion that the FIE occurs only when recognizing faces already stored in memory, this conclusion is surprising given the widely held view that the "face inversion effect is really a perceptual phenomenon rather than a memory phenomenon" (Freire, Lee & Symons 2000; p.160). An alternate explanation more compatible with this view would be possible if participants can strategically use the results of configural processing. That is, if configural encoding is not an automatic process but rather that participants will utilize it only when they know it will improve performance for all items. For example, if an item was encoded using purely featural information (e.g., studied inverted) it might be detrimental to use configural information at test, as suggested by the encoding specificity effect; if only featural information is available from study, performance would benefit from a matched (featural) test encoding but hurt by a mismatched (configural) encoding.

Participants in the *Test Inverted* condition may have relied completely on featural information because configural information at test was not available (or was too difficult to extract) from the inverted test traces. In contrast, participants in the *Test Upright* condition had configural information available but may have instead relied on featural information because they had no way of knowing for which items the configural information would be detrimental (i.e., those studied inverted). In future research we will test this possibility by modifying testing in the *Test Upright* condition so participants are given separate test lists in which they are informed of an item's study orientation (assuming it is old). In this situation, the use of configural information for test items studied upright could be beneficial, in which case the strategic hypothesis predicts a non-monotonic state-trace plot.

Acknowledgments

Thanks to Jessica Barton, Adam Beaman, Therese Cubis, Minya Griffiths, Alison Hodgson, Marc Inberg, Sarah Jemmett, Sarah Johns, Cindy Kuzarevski, Simon Merriman, Daniel Milford, Namrata Murti, Belinda Preston, Victoria Todd and Mitchell Wall for assistance with stimulus preparation and running participants.

References

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19,* 137, 137-181.

Barton, J.J.S., Keenan, J.P., & Bass, T. (2001). Discrimination of spatial relations an features in faces: Effects of inversion and viewing duration. *British Journal of Psychology, 92,* 527-549.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science, 195,* 312-314.

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

Dunn, J.C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95,* 91-101.

Freire, A., Lee, K., & Symons, L.A. (2000). The face inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception, 29,* 159-170.

Hancock, P.J.B., Bruce, V., & Burtong, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Science, 4,* 330-337.

Leder, H. (1999). Matching person identity from facial line drawings. *Perception, 28,* 1171-1175.

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

Loftus, G.R. (1978). On interpretation of interactions. *Memory & Cognition, 6,* 312-319.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476-490.

Loftus, G.R., Oberg, M.A., & Dillon, A.M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review, 111,* 835-863.

Maurer, D., LeGrand, R., & Mondloch, C.J. (2002). The many faces of confdigural processing. *Trends in Cognitive Science, 6,* 255-260.

Newell, B.R., & Dunn, J.C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences, 12,* 285-290.

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

Phillips, P.J., Wechsler, H., Huang, J., & Rauss, P.J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing, 16,* 295-306.

Rakover, S.S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology, 93,* 1-30.

Rakover, S.S., & Teucher, B. (1997). Facial inversion effects: Parts and whole relationship. *Perception & Psychophysics, 59,* 752-761.

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

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

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

Valentine, T., & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61,* 259-273.

Yarmey, A.D. (1971). Recognition memory for familiar "public" faces: Effects of orientation and delay. *Psychonomic Science, 24,* 286-288.

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