Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Disproportionate Face Inversion Effect in Recognition Memory

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Prince, Melissa
Heathcote, Andrew

Publication Date
2010

Peer reviewed

The Disproportionate Face Inversion Effect in Recognition Memory

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

Abstract

The Disproportionate Face Inversion Effect (DFIE), the finding that inversion disproportionately affects face recognition, provides a primary piece of evidence to suggest that faces are processed in a qualitatively different way to other visual stimuli (i.e., along configural as well as featural dimensions). However, when Loftus, Oberg and Dillon (2004; also Prince and Heathcote, 2009) examined the DFIE 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 task. We further examine this one dimensional result with more precise individual measurement and more specifically, Prince and Heathcote's suggestion that the use of configural encoding may not be automatic in recognition memory.

Keywords: Disproportionate Face Inversion Effect; Recognition Memory; State-trace analysis.

Over the course of a human lifetime, thousands of faces can become so familiar that they can be recognized after only a glance, when seen in an unfamiliar context and even after undergoing significant physical changes (Maurer, LeGrand & Mondloch, 2002). Indeed, the common experience of recognizing a familiar face in a crowd or involuntarily imagining a face in scenic features seems to indicate that humans possess an innate aptitude for face processing. However, this expertise is less evident when the faces are unfamiliar (Hancock, Bruce & Burton, 2000), and even more so when they are viewed upside-down.

It is widely found that perception and memory performance for all mono-oriented stimuli (i.e., objects usually viewed in a specific "upright" orientation) are strongly disadvantaged by inversion; called the *Inversion Effect.* However, in his seminal paper, Yin (1969) observed that this inversion effect was disproportionately stronger for faces compared to mono-oriented control stimuli (e.g., houses) that were matched as closely as possible to faces in terms of familiarity, complexity and difficulty in applying a verbal label; known as the *Disproportionate Face Inversion Effect* (DFIE). Although the inversion effect is taken to indicate there is a general factor affecting the processing of all mono-oriented stimuli, the DFIE suggests there is an additional face specific factor. Hence the DFIE has become one of the primary pieces of evidence to suggest that face processing is "special".

In this paper, we aim to explore the evidence for the DFIE in recognition memory accuracy for unfamiliar faces. In particular, we will focus on an alternate statistical method for testing the effect of inversion called *State-Trace Analysis* (Bamber, 1979). Using this technique, Loftus, Oberg and Dillon (2004) found that, in contrast to results from traditional analyses that revealed a weak DFIE, state-trace results indicated that unfamiliar faces were not special relative to other mono-oriented stimuli when inversion was only manipulated during the encoding stage of a recognition memory task. Loftus et al. therefore suggested that the DFIE only occurs during memory retrieval. Although Prince and Heathcote (2009) replicated this state-trace result, as well as ruling out several potential caveats on Loftus et al.'s methodology and state-trace analyses, they questioned the memory retrieval interpretation. Here we examine an alternate explanation for these results, namely Prince and Heathcote's *Strategic Hypothesis.*

The Disproportionate Face Inversion Effect

Since Yin's (1969) initial demonstration, the DFIE in recognition memory has been replicated numerous times and with various procedural variations. Although many studies have followed Yin's original design where items were studied upright or inverted and tested in the same "matched" orientation, a DFIE has also been found when images were tested using a different viewpoint from study (Valentine & Bruce, 1986) as well as when all images were studied upright but tested upright or inverted (Yarmey, 1971). Consequently, the DFIE has been taken to indicate that face processing is qualitatively different from the processing of other visual stimuli.

It has been suggested that the two factors (or dimensions) underlying the DFIE might be two types of information that can be extracted from the images. The first, *featural information*, is common to all mono-oriented stimuli and refers to the isolated features of an object that can be specified without reference to its other parts. In contrast, the second type, *configural information,* is mostly or only available to faces and enables good discrimination despite the highly similar structure and features that all faces share (McKone & Yovel, 2009). At least three types of configural information have been proposed: (a) *holistic information,* which captures the overall look of a face; (b) *first order relational information,* which refers to the arrangements of features that define a face; and (c) *second order relational information,* which refers to distances between internal features. However, the differences between these sub-types are not of critical importance here. Rather, what is important is the general finding that inversion differentially affects two broad classes of largely independent information.

Although both featural and configural information are affected by inversion, it is typically found that the extraction of configural information is particularly disrupted. Hence, it is often believed that upright faces are processed using both featural and configural information, whereas only featural information is available for inverted faces. Recent evidence, however, suggests a more graded relationship, such that inversion decreases the rate at which both featural and configural information can be extracted from a face, but to a greater degree for configural information (Valentine & Bruce, 1986).

Identifying Dimensions of the DFIE

Evidence for the DFIE, and hence for the existence of two underlying dimensions for face processing, is traditionally provided by a dissociation quantified by an interaction test comparing the size of the inversion effect for faces (the *face inversion effect*; FIE) to that for a mono-oriented control stimulus, such as houses (the *house inversion effect*; HIE). However, it has been argued that such dissociation logic at best makes the rejection of a one-dimensional account more plausible or parsimonious. Moreover, where response measures are bounded (e.g., accuracy data), interactions may be scale dependent (e.g., influenced by floor and ceiling effects; Loftus, 1978). An alternate method proposed to overcome the caveats on dissociation logic is *State-trace analysis* (e.g., Newell & Dunn, 2008). State-trace analysis provides a rigorous method for determining whether a single dimension (i.e., a single latent variable) is able to explain the joint effect of two or more experimental factors, and assumes only that the mapping between the latent variable and response is monotonic (i.e., that the response and latent variable consistently change in the same direction).

The results from state-trace analysis are assessed using a *state-trace plot*, which is essentially a scatterplot showing the covariation of two factors, namely the *state* and *dimension* factors. As shown in Figure 1, the state factor defines the axes of this plot, while each level of the dimension factor typically defines a set of points within the plot. In particular, the dimension factor is manipulated with the aim of differentially influencing the latent variables. In applications examining the DFIE, the state factor is defined by recognition accuracy for face and house images and the dimension factor manipulated to differentially influence the latent configural dimension is the image orientation.

The crucial diagnostic feature of this plot concerns whether or not the data fall on a single monotonic function; that is, whether the ordering of the x-axis values is the same as the ordering of the y-axis values. At least three data points are required to potentially violate monotonicity, and thus a third factor, called the *trace factor*, is introduced to sweep out a set of points (i.e., a "data trace") within each level of the dimension factor. Importantly, the trace factor must itself have a monotonic effect if we are to unambiguously attribute dimensionality evidence to the interaction between the state and dimension factors (i.e., that $A<B$ and $a<b$ in Figure 1). Loftus et al. (2004), for example, manipulated the study presentation time, which can reasonably be assumed to have a monotonic effect on accuracy; shorter study durations always lead to poor recognition in all conditions (within measurement limits).

If the two levels of the state factor depend on the same underlying dimension, the points on a state-trace plot will fall on a single monotonic function (e.g., in Figure 1a the x- and y-axis order is a,A,b,B). If, however, performance for each state is determined by more than one dimension (e.g., along featural and configural dimensions), the resulting state-trace plot will be non-monotonic (see Figure 1b). It is important to note that although a non-monotonic plot cannot have been produced by a one-dimensional model, the converse does not necessarily hold. Monotonic evidence is only diagnostic of dimensionality when there is overlap of the data traces on at least one axis. Where data-trace overlap fails (such as in Figure 1c), a state-trace plot can be monotonic even if two separate dimensions exist.

Despite the simplicity of state-trace analysis graphically, the best statistical method for testing departures from monotonicity remains an open question (e.g., Newell & Dunn, 2008). Recently Heathcote, Brown and Prince (submitted) proposed a method for assessing dimensionality in state-trace designs based on a Bayes Factor method of selecting amongst models defined by ordinal constraints: namely, (a) a *non-trace* (NT) model, which assumes the trace factor does not have a monotonic effect on performance: that is that the trace model is violated; (b) a *no overlap* (NO) model, which given the trace model holds, assumes the data traces do not overlap and hence cannot be

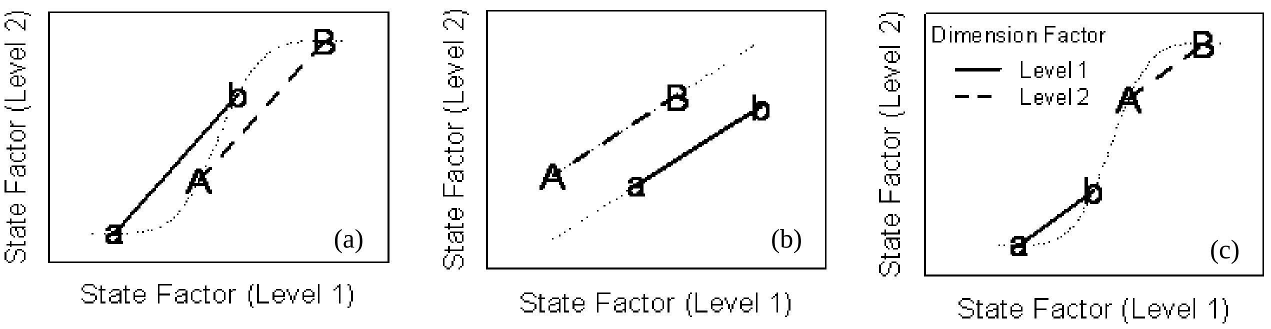

Figure 1: Example state-trace plots for a design where the state, dimension and trace factors each have two levels. The thin dotted lines show the underlying dimension or processes revealed in the plot, with examples of (a) a one-dimensional plot, (b) a two-dimensional plot and (c) a non-diagnostic state-trace plot (i.e., due to the data traces having no overlap)

considered diagnostic of dimensionality; (c) a *uni-dimensional* (UD) model, which assumes the state-trace plot is monotonic, given that both trace monotonicity and data trace overlap hold; and (d) a *multi-dimensional* (MD) model, which assumes the state-trace plot is non-monotonic, again given that both trace monotonicity and data trace overlap hold. Together these four models account for all possible orders. The aim is then to find the model with the highest posterior probability; that is, the model with the highest probability of being the data generating model.

A Memory Retrieval Phenomenon?

Using state-trace analysis, Loftus et al. (2004) reported an apparent exception to the otherwise robust DFIE result; evidence for a single dimension in accuracy averaged over participants in recognition memory for unfamiliar faces (Experiment 1). In contrast, when the faces were famous (i.e., familiar; Experiment 2), they found evidence for more than one dimension. In both experiments, images were studied upright or inverted and all tested upright. This design was utilized to examine Valentine's (1988) assertion that "the orientation of the inspection [study] series does not appear to be critical" (p.474) to produce a DFIE. Loftus et al. concluded that a DFIE would only emerge when inversion was present at the time of memory retrieval, because familiar faces cause memory retrieval at study (and so produce a DFIE when inversion occurs during study), but unfamiliar faces do not (so inversion occurring at study cannot cause a DFIE).

Although, Prince and Heathcote (2009) replicated the one-dimensional state-trace result with unfamiliar faces, the conclusion that a DFIE only occurs at memory retrieval goes against the general opinion in the literature which would suggest the DFIE "is really a perceptual phenomenon rather than a memory phenomenon" (Freire, Lee & Symons, 2000; p.160). Consequently, Prince and Heathcote proposed an alternate explanation more compatible with this perceptual view, whereby participants may be able to strategically use configural information, but only when they know it will improve performance for all items. That is, that the use of configural information may not be automatic in recognition memory.

Here we aim to further examine the one-dimensional state-trace result for unfamiliar faces, as well as Prince and Heathcote's (2009) strategic hypothesis. To do so we ran new experiments that greatly increased the number of observations obtained from each participant (78 observations per design cell), by increasing the number of trials and reducing the number of study durations. Our first condition partly replicated Prince and Heathcote's *Test Upright* design, with both upright and inverted study trials mixed in each study list and all items tested upright. However, it used a two-alternative forced choice (2AFC) recognition memory test, rather than the single-item testing used in the original study (i.e. on each test trial participants chose between a studied and unstudied face, or between a studied and unstudied house). This condition was run to check if Prince and Heathcote's result was replicable with a different testing procedure and with a slightly different, and more powerful, design. We denote this condition *TUM2* (*T*est *U*pright, *M*ixed study lists, *2*AFC).

In *TUM2* (thus also Prince and Heathcote's, 2009, *Test Upright* design), an old item can either be studied and tested upright or studied inverted and tested upright. The former case has a matched (configural) encoding available at study and retrieval. However, when an image is studied inverted it only (or at least mostly) can be encoded using featural information, yet configural information is available from the upright test presentation. As suggested by the encoding specificity effect (i.e., the improvement in memory when study and test conditions match; Tulving & Thomson, 1973), if only featural information was available at study, performance would benefit from a matched (featural) test encoding and be hurt by a mismatched (configural) encoding. Hence it may be detrimental for participants to use configural information when an item had been studied inverted.

In these upright test conditions, upright and inverted items were mixed together at study. Therefore, when all items are presented upright at test, participants have no way of knowing for which test items the use of configural information may be detrimental (i.e., those studied inverted). As these experiments used multiple study-test cycles participants would quickly become aware that all test items were upright. Hence it is possible that they decided to rely purely on featural information, either by not encoding upright study items along a configural dimension, or choosing not to use the configural information available at test. In either case, both faces and houses would only be encoded along a single (featural) dimension, producing the one-dimensional state-trace plots observed by Loftus et al. (2004) and Prince and Heathcote (2009).

To test this possibility, in our second condition, participants viewed two types of study-test lists where: (a) all items were studied and tested upright or (b) all items were studied inverted and tested upright. By blocking study orientation in this manner we hoped that participants would become aware of when configural encoding was advantageous (in type 'a' lists) and hence make use of it. If this occurred, we should observe a multi-dimensional state-trace plot, and hence evidence against Loftus et al.'s (2004) memory retrieval hypothesis. We denote this condition *TUB2* (*T*est *U*pright, *B*locked study lists, *2*AFC).

Method

Participants

The 38 participants were recruited from members of the wider community, who had normal or corrected-to-normal vision. They received cash reimbursement for their time (total AUD$30.00). Two subjects in TUM2 were excluded due to their raw percentage correct falling below 55%, leaving 18 subjects in *TUM2* and *TUB2*.

Stimuli

Stimuli were black and white bitmap images (120 x 105 pixels) displayed at twice their original size. A total of 936 face stimuli were sourced from the FERET database (Phillips, Wechsler, Huang & Rauss, 1998), excluding images with averted gaze, distinctive facial expressions or blemishes (either natural or the result of photographic process). These face stimuli were divided into homogenous blocks based on race, gender and any other distinctive feature (i.e., glasses or facial hair). An additional 36 Caucasian males without facial hair or glasses were included for the practice phase.

A total of 936 house stimuli (with an additional 36 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 largely drawn from this region. Following Prince and Heathcote (2009), house stimuli were also divided into homogenous blocks based on their most distinctive feature (e.g., fence, two-storey).

Apparatus

Testing was completed either at individual computer terminals equipped with 17inch LCD monitors or at an external location 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

It was emphasized during the instructions for the task, that the orientation of a stimulus was irrelevant to a recognition decision; that is, participants should identify an image as being “old” even if the test item had been studied in a different orientation. In *TUB2*, participants were further informed that study lists would be comprised of either all upright or all inverted images and a warning was displayed prior to each study list indicating the study orientation to be used. Before commencing the main experiment, participants completed two full length practice blocks; one for faces and one for houses, with order counterbalanced over participants.

A study list (comprised of 18 trials) was initiated by pressing the space bar, following which the warning “*Prepare for study … of … Place your fingers on the keys*” was 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 (upright: 33, 100, 267ms; inverted: 267, 800, 2048ms), with durations selected to maximize data trace overlap and each duration level used equally often in every study list. After each study presentation, 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 from “definitely no” to “definitely yes”. The purpose of this prospective confidence judgment was to encourage participants to attend to the stimulus and this data will not be considered further.

The test list (again comprised of 18 trials) was marked by a 300ms blank screen, followed by the warning “*Prepare for test … of … Place your fingers on the keys*” displayed for 2000ms. Each test trial was preceded by a blank screen following which the test item and retrospective confidence response scale were presented for a maximum of 5000ms. For our 2AFC design, a pair of test images (one old and one new, with the old item appearing equally often on the left and right) were presented above the question “*Which image was previously studied and how confident are you that you have seen this image earlier?*” Again participants responded using a four-point scale from “definitely left” to “definitely right”. For the entire length of the study and test lists, the words “STUDY” and “TEST” were respectively displayed in the top left corner of the screen.

Following the practice study-test lists, participants received feedback on the number of times they used each of the confidence levels. The purpose of this feedback was to encourage participants to use the full range of the confidence scale.

Participants were required to attend three one hour sessions, preferably on consecutive days. Participants completed 12 study-test lists in their first session and 20 study-test lists in the later two. At the end of each list participants were able to take a self paced break (minimum of 30s), while three longer breaks (minimum of 5min) occurred within each one hour session.

Results

The retrospective confidence rating was used to determine a participant’s probability correct (i.e., the number of trials correct divided by total number of trials). Accuracy was then quantified by the inverse cumulative normal probability (z) transformation of the probability correct.

We first report a preliminary analysis to ensure the present study was able to replicate previous findings that accuracy is linear as a function of the logarithm of study duration. One-way repeated measures ANOVAs were performed on the effect of the logarithm of study duration for upright and inverted houses and faces in each condition with polynomial trend analyses. Linear trends were all statistically reliable ($p<.05$) and accounted for almost all (minimum 88%) of the variance in accuracy as a function of study duration. The only quadratic trends to approach significance were for *TUM2*’s upright faces ($p=.045$) and upright houses ($p=.075$).

Evidence for the DFIE was first assessed by the traditional test of an interaction between orientation and stimulus type. As the 267ms duration level was the only study duration common to both upright and inverted stimuli, the DFIE was tested by a two-way (orientation by stimulus type) ANOVA using only the 267ms data. Table 1 also shows estimates of the inversion effect (i.e., the difference between upright and inverted) for faces (FIE) and houses

(HIE), the corresponding DFIE estimates (DFIE=FIE-HIE) and the results of associated *t*-tests.

Table 1: Estimates of the FIE, HIE, and DFIE and results associated t-tests, for the 267ms data.

	FIE	HIE	DFIE
TUM2	0.281**	0.270***	0.011
TUB2	0.274***	0.215***	0.059

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

For *TUM2* there was no reliable difference in accuracy between houses (*M*=0.479) and faces (*M*=0.411), *p*=.152. However, accuracy was reliably higher for upright items (*M*=0.582) than inverted items (*M*=0.307), *F*(1,17)=30.50, *p*<.001. Although a slightly greater inversion effect was observed for faces than houses (DFIE=0.01), this effect was not statistically reliable, *p*=.91. Similarly for *TUB2*, accuracy was higher for houses (*M*=0.421) than faces (*M*=0.409), but not reliably so, *p*=.83. Upright items were again reliably more accurate (*M*=0.537) than inverted items (*M*=0.293), *F*(1,17)=41.93, *p*<.001. However, there was no reliable DFIE (DFIE=0.059; *p*=.42).

State-trace plots for each condition are shown in Figure 2. Results for upright study are joined, as are the points for inverted study. These lines are clearly monotonically increasing, and consistent with the requirement that the trace factor has a monotonic effect, both conditions' posterior model probabilities favored the trace model being true, p(NT)<.001. The plots also show excellent data trace; for both *TUM2* and *TUB2* p(NO)<.001. In assessing the overall dimensionality, *TUM2* showed positive evidence for a multi-dimensional model, p(MD)=0.910, however, TUB2 showed equivocal evidence suggestive of a one-dimensional account, p(UD)=0.733.

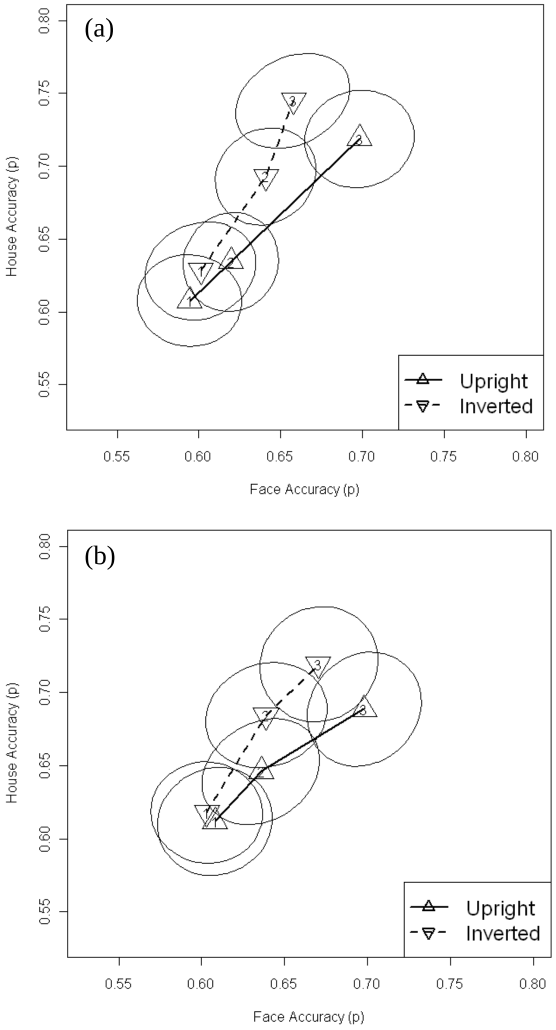

Figure 2: State-trace plots showing the 50% credible regions for the (a) *TUM2* and (b) *TUB2* conditions. The numbers 1…3 indicate shorter to longer study durations.

Discussion

We replicated Loftus et al.'s (2004) and Prince and Heathcote's (2009) finding of a linear increase in accuracy consistent with the suggestion that there was no abrupt change in strategy (i.e., no switch from featural to configural processing) associated with longer study durations. Additionally, we replicated the lack of evidence for a DFIE using the traditional interaction measure (although the DFIE estimates were of the same magnitude as Loftus et al., and Prince & Heathcote). Our state-trace findings, however, were mixed.

We found clear multi-dimensional evidence consistent with the use of both featural and configural information for *TUM2*, where inversion was only manipulated during initial encoding and upright and inverted items were mixed together at study. However, for *TUB2*, where study lists were blocked by orientation, we observed evidence suggestive of a single underlying dimension (although at an equivocal level). In this blocked condition, participants were informed of an item's study orientation if it was old and therefore, according to Prince and Heathcote's (2009) strategic hypothesis, may have been able to reinstate the use of configural information. The observed one-dimensional result, however, does not offer support for this proposal. This is not to say that our results are consistent with Loftus et al.'s (2004) memory retrieval hypothesis. Indeed the strong multi-dimensional result for *TUM2* cannot be explained by a memory retrieval interpretation, as orientation was only manipulated during initial encoding.

It is important to note that the posterior model probabilities on which we are basing our interpretations, are not simply the average result over participants. Rather they examine, for example, the probability that *all* individual state-trace results are one-dimensional versus *all* being multi-dimensional. Hence these probabilities can sometimes be influenced by outlying subjects. To ensure our results were not influenced in this way, we re-examined the dimensionality results, excluding participants with poor evidence (*p*>.5) for trace monotonicity and data trace overlap (four participants from *TUM2* and seven from *TUB2* were excluded using this criteria). However, both conditions revealed the same pattern of results; that is, multi-dimensional evidence for *TUM2* and equivocal evidence for *TUB2*. Although *TUB2* showed a decrease in the probability supporting a one-dimensional model, p(UD)=0.691.

One possible explanation for observing multi-dimensional evidence, even though inversion was only manipulated during the initial stimulus encoding, is that our more precise individual measurement also produced higher accuracy performance overall and consequently an improved effect size. Although state-trace analysis is not affected by floor and ceiling effects to the same degree as traditional dissociation analyses, if accuracy is not high enough to reveal the decrement caused by inversion then it will also not be able to reveal the underlying dimensionality. Consistent with this suggestion, we can observe from the *TUM2* state-trace plot that the data traces do not depart from monotonicity (indicating multi-dimensional evidence) until the longer study duration levels (where accuracy is also higher). This same pattern can also be seen to a lesser degree in *TUB2.*

Although not reported here, we also ran these same mixed and blocked conditions using a yes/no testing procedure (i.e., participants were shown a single test item and asked to indicate if that item had or had not been studied), and in contrast to our 2AFC results, observed equivocal one-dimensional evidence. Interestingly, it has been found that memory performance is advantaged by a 2AFC procedure over a yes/no procedure (Deffenbacher, Leu & Brown, 1981), which could explain why our 2AFC conditions tended toward multi-dimensional evidence. It should also be noted that recognition memory studies in general tend to show a smaller inversion effect than perceptually based studies (e.g., *TUM2* showed a 9.97% drop in accuracy, but perceptual tasks can show a drop double this magnitude; see McKone & Yovel, 2009). Hence evidence for more than one dimension underlying face processing may only emerge when performance is high enough to reveal the decrement caused by stimulus inversion.

We will pursue two avenues in future research. First, as our results did not offer strong insight into Loftus et al.'s (2004) memory retrieval hypothesis we will examine state-trace evidence for the DFIE in recognition memory using a paradigm in which unfamiliar faces are all studied upright and tested either upright or inverted. In this paradigm, Loftus et al.'s (2004) memory hypothesis predicts that evidence for multiple dimensions should emerge, because inversion occurs at test where memory retrieval is required. Second, we will extend the use of state-trace analysis to a perceptual paradigm, such as a sequential same-different task, in order to investigate whether evidence for more than one dimension emerges with larger inversion effects.

Acknowledgments

Thanks to students enrolled in Psyc4815/Psyc6781 at the University of Newcastle, 2009 for assistance with stimulus preparation and running of participants.

References

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

Deffenbacher, K.A., Leu, J.R., & Brown, E.L. (1981). Memory for faces: Testing method, encoding strategy, & confidence. *American Journal of Psychology, 94,* 13-26.

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., & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Science, 4,* 330-337.

Heathcote, A., Brown, S. & Prince, M. (submitted). The design and analysis of state-trace experiments.

Loftus, G.R. (1978). On interpretation of interactions. *Memory & Cognition, 6,* 312-356.

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

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review, 16,* 778-797.

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

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.

Prince, M., & Heathcote, A. (2009). State-trace analysis of the face-inversion effect. In N.A. Taagen & H. van Rijn (Eds.). *Proceedings of the 31st Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

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. *Acta Psychologica, 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.