LEARNING TO PERCEIVE ORIENTATION STRUCTURE

PERCEPTUAL LEARNING OF THE ORIENTATION STRUCTURE OF FACES AND TEXTURES

By ALI HASHEMI, B.Sc. (Honours)

A Thesis Submitted to the Department of Psychology, Neuroscience & Behaviour, and the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2018) Hamilton, Ontario,
(Psychology, Neuroscience & Behaviour)

TITLE: Perceptual learning of the orientation structure of faces and textures
AUTHOR: Ali Hashemi, B.Sc. (Honours in Biology & Psychology, McMaster University)
SUPERVISORS: Dr. Allison B. Sekuler & Dr. Patrick J. Bennett
NUMBER OF PAGES: xii, 199

Abstract

Perceptual learning occurs because observers become more sensitive to informative aspects of the stimuli. Learning the informative aspects of one stimulus set does not transfer to another stimulus set of the same class. In this dissertation, the argument will be made that if observers learn how to *discover* informative aspects, learning will be more generalizable. However, discovery requires that the informative aspects are not easily apparent. To this end, stimulus orientation structure can be manipulated to contain informative structure in one orientation band, and non-informative structure in the other orientation band. Such a manipulation was inspired by research on face perception: Faces are best identified when decisions are based more on the horizontal relative to the vertical facial structure. Hence, the first three chapters focus on understanding the horizontal bias during face identification, and the final two chapters introduce a novel stimulus set for which horizontal bias may be learned. Chapter 2 identifies a neural marker of horizontal bias that is correlated with face identification accuracy, suggesting that we can predict how well observers identify faces based on their neural sensitivity to horizontal relative to vertical structure. Chapter 3 shows that when face identification accuracy declines due to healthy ageing, so too do behavioural and neural horizontal bias, but Chapter 4 shows that perceptual learning can increase horizontal bias in healthy older adults. Chapter 5 uses texture stimuli and shows that observers can learn to discover informative horizontal structure embedded in uninformative vertical structure. Chapter 6 extends these findings to show that adequate practice results in learning that generalizes to novel textures for which the orientation-selective processing is relevant. The results presented inform our understanding of the neural representations associated with orientation-selective processing, and suggest that observers can learn to discover informative structure conveyed by a particular orientation band.

Acknowledgements

To my family, friends, and colleagues,

Thank you first and foremost to my supervisors, Allison Sekuler and Patrick Bennett, for shaping me into the scientist I have become. I leave your lab certain that I've learned how to think critically, ask good questions, and reach success. Allison, thank you for being the first professor to reply to my email seeking an undergraduate thesis supervisor – I had no idea that my entire career path would change with that email. Pat, thank you for showing me that statistics are simple, and that simple statistics are best (N.B. linear contrasts). Thank you also to Louis Schmidt for providing sound advice and constantly supporting all of my projects and at every committee meeting.

Everyone I met in PNB, thank you for making it easy to enjoy my time as a graduate student in the lab and department, and for that one very nerdy week every year in Florida. A special mention to Donna Waxman, Matthew Pachai, Jessica Cali, Bruce Milliken, and Justin Balsor, all of whom I value as my collaborators, mentors, and friends. Donna, absolutely none of my work could be accomplished without everything you do, and I am so thrilled to have made a lifelong friend in you. Matt, you took me under your wing when I was an undergraduate, and in the dozen experiments together, your ideas pushed me to be a better scientist. Jessica, your skepticism and friendship are unique and of the highest quality. Bruce, your view of the world is inspiring, and hearing your enthusiasm for the big picture has improved my way of thinking. Justin, having my best friend around the corner, literally, made every day better.

None of this would be possible without my parents' decision to come to Canada for our education. Mom and dad, thank you for the daily sacrifices, regular phone calls, delicious home cooked meals, and belief in my pursuit for a career in reading, writing, and thinking about incredibly specific things. Who would have thought that when I would still be at McMaster ten years later? Lastly, and most importantly, thank you to my dear wife. Nicole, you've supported me through this journey in so many ways, and I am forever grateful that I can come home to your unlimited love.

Thank you.

Ali

Contents

List of Figures

Declaration of Academic Achievement

Chapter 1: General introduction
Author: Ali Hashemi

Chapter 2: The role of horizontal facial structure on the N170 and N250
Authors: Ali Hashemi, Matthew V. Pachai, Patrick J. Bennett, and Allison B. Sekuler
Contributions: AH, MVP, PJB, and ABS conceived and designed the study. AH conducted the analysis. AH wrote the manuscript with revisions from MVP, PJB, and ABS. Reprinted, with permission, from the manuscript published in *Vision Research*.

Chapter 3: The behavioural and neural horizontal bias of older adults
Authors: Ali Hashemi, Patrick J. Bennett, and Allison B. Sekuler
Contributions: AH conceived and designed the study with suggestions from PJB & ABS. AH conducted the analysis. AH wrote the manuscript with revisions from PJB and ABS.

Chapter 4: Practice identifying faces recovers horizontal bias in older adults
Authors: Ali Hashemi, Patrick J. Bennett, and Allison B. Sekuler
Contributions: AH conceived and designed the study with suggestions from PJB & ABS. AH conducted the analysis. AH wrote the manuscript with revisions from PJB and ABS.

Chapter 5: Learning to discover orientation-specific structure in textures
Authors: Ali Hashemi, Matthew V. Pachai, Allison B. Sekuler, and Patrick J. Bennett
Contributions: AH conceived and designed the study with suggestions from MVP. AH

conducted the analysis with suggestions from PJB. AH wrote the manuscript with revisions from ABS and PJB.

Chapter 6: Extended training of orientation filtered textures increases generalization of learning

Authors: Ali Hashemi, Allison B. Sekuler, and Patrick J. Bennett

Contributions: AH conceived and designed the study. AH conducted the analysis. AH wrote the manuscript with revisions from ABS and PJB.

Chapter 7: General discussion

Author: Ali Hashemi

Chapter 1

General Introduction

1.1 Perceptual Learning

Normal perceptual development requires adequate sensory experience. Visual development is characterized by the existence of sensitive periods, during which times the development of the visual system is particularly sensitive to the influence of visual experience (Lewis and Maurer, 2005; Siu and Murphy, 2018). For example, the development of normal stereopsis requires normal binocular visual experience before 6 months of age (Birch and Petrig, 1996), whereas normal motion perception is sensitive to visual experience until up to 8 years of age (Hadad et al., 2015). Importantly, although visual developmental is most responsive to experience during sensitive periods, changes in visual perception do occur throughout the lifespan (Karni and Bertini, 1997). For example, changes in vision and other sensory modalities can occur as a result of perceptual learning: such changes are often long-lasting (Karni and Sagi, 1993) and are thought to reflect changes in the properties of sensory neurons (for a review see Sagi, 2011).

One of the hallmarks of perceptual learning is that performance improvements often are stimulus-specific. In vision, perceptual learning is specific to stimulus orientation, retinal position, and sometimes even the input eye (e.g., Karni and Sagi, 1991). Stimulus-specificity has been found in several tasks: for example, practice discriminating a vertical vernier stimulus does not transfer to a horizontal stimulus (McKee and Westheimer, 1978); practice discriminating the orientation or spatial frequency of gratings does not transfer to untrained orientations and spatial frequencies

(Fiorentini and Berardi, 1981); and practice discriminating the motion direction of moving dots does not transfer to novel directions (Ball and Sekuler, 1987). Stimulus-specific perceptual learning has also been demonstrated when discriminating complex textures (Hussain et al., 2009a) and faces (Hussain et al., 2009b), which are of particular interest to this thesis.

In a series of studies, Hussain and colleagues characterize the perceptual learning of textures and faces. Practice in a one-of-ten texture identification task improves accuracy, but the effect of practice does not transfer to novel textures, to trained textures presented at a novel orientation, or to contrast-reversed versions of the trained textures (Hussain et al., 2009a). Small amounts of practice are required to observe the effects of perceptual learning on texture and face identification (Hussain et al., 2009b), and from the outset of training, the learning is specific for the trained textures (Hashemi et al., 2013) and faces (Hussain et al., 2012b). Stimulus-specific perceptual learning of textures persists for at least one year after training (Hussain et al., 2011). Gold et al. (2004) showed that perceptual learning in a texture discrimination task reflected observers basing their decisions on more informative aspects of the stimuli. Observers learned to discriminate textures by extracting diagnostic information from specific regions of the stimuli. Interestingly, the particular regions differed between observers, indicating that there was no spatial region that observers universally found diagnostic across the set of textures. Face discrimination, which is also affected by perceptual learning (Gauthier et al., 2003; Gold et al., 1999), is based on diagnostic information conveyed by spatial regions around the eyes and brows (Sekuler et al., 2004; Peterson and Eckstein, 2012). Unlike with textures, the particular regions used to discriminate faces are similar across healthy adults (e.g., Sekuler et al., 2004; Gosselin and Schyns, 2001), which is likely due to fact that the spatial features in faces, unlike filtered textures, are arranged similarly.

The findings from studies of the perceptual learning of textures and faces suggest that the probability that observers will adopt a particular strategy is, naturally, dependent on the distribution of the diagnostic information in stimuli. Thus, the structure of the stimulus may play a significant role in the stimulus-specificity of learning. For example, in a texture identification task, sampling from a subset of informative regions is an effective strategy, but only for the textures in the trained set. Given that textures used in those studies were generated from noise, there is no pre-determined configuration of the stimulus aspects that can be predicted in novel

items, and therefore learning to identify six textures based on information sampled from several specific spatial regions is unlikely to transfer to a different set of six textures. In other words, the stimulus-specific nature of perceptual learning is consistent with the idea that practice enables observers to identify textures based on diagnostic features of the particular textures seen during training, rather than base identification on more general processes that can be applied to the entire class of textures seen during training. Hussain et al. (2012a) showed that if novel stimuli are shown on every trial, texture identification accuracy still improved significantly with practice, albeit at a slower rate than when the textures remained constant during training. Since it was impossible to learn particular aspects tied to any one exemplar, observers must have learned aspects that apply to the general class of textures, and therefore were able to generalize learning to novel textures. On the one hand, observers can learn specific idiosyncrasies of a few stimuli, and on the other hand observers can learn idiosyncrasies that apply to an entire class of stimuli by being exposed to a significant proportion of stimuli that belong to that class. Between these two extremes might exist a situation in which observers learn to *discover* what stimulus characteristics are diagnostic. In such a circumstance, it is possible that the observer may discover and therefore base decisions on which visual channels carry diagnostic and non-diagnostic information.

1.2 Learning through discovery

The idea of learning to discover the diagnostic information resembles how expertise is acquired by those specialized in using medical imaging to detect often subtle signs of a disease or dysfunction (e.g., Norman et al., 1989; Nodine et al., 1999). Experts often can classify medical images taken from healthy and non-healthy patients in a single glance (Drew et al., 2013), reflecting better detection and processing of diagnostic aspects of the image (Sowden et al., 2000). Interestingly, experienced radiologists have higher sensitivity to the low-contrast features of x-ray images, but training novices to detect these low-contrast features does not result in a better ability to classify x-ray images (Sowden et al., 2000). This asymmetry in the transfer of perceptual learning suggests that increased sensitivity to low-contrast features in x-rays is not a sufficient explanation of the abilities of expert radiologists.

In the identification study by Hussain et al. (2012a), the diagnostic information could not have been the structure of a particular texture, since no texture was repeated

twice. Therefore, other than detecting statistical regularities within the entire class of stimuli (Sagi, 2011), the exact strategy that observers employed, and improved on, is unknown. Since stimulus structure, and how stimulus structure varies across trials, influence what is learned with practice, then the stimulus may be manipulated to promote a desired strategy. For instance, much like medical imaging, diagnostic information can be embedded in an uninformative context in such a way that the diagnostic information is not immediately noticeable to the observer. With such stimuli, it may be possible to train observers with a small set of stimuli, but make it less likely that they learn idiosyncrasies that are tied to the small set of stimuli, and therefore be more likely to generalize learning to new stimuli.

One way to modify texture stimuli to *hide* the diagnostic information is by placing informative and non-informative structure in different orientation bands. For instance, the experiments in Chapters 5 and 6 of this thesis used a set of six textures that were filtered to contain task-relevant structure only in a horizontal orientation band, while the orthogonal orientation band contained task-irrelevant structure derived from a texture that was not in the set of six textures. The resulting stimuli appear to be textures with structure in all orientations, but unbeknownst to the observer only horizontally oriented structure is informative for the texture identification task. The working hypothesis is that instead of learning to *discriminate* structure, observers will learn to *discover* discriminable/relevant structure. Learning to discover may be less specific to particular aspects about the trained stimuli, and therefore generalize to novel textures where the learned strategy is relevant. Learning to discover task-relevant information is similar to the idea that perceptual learning occurs in a central location of the processing pathway, and that specificity to stimulus characteristics is not due to *where* learning occurs along the visual pathway, but rather *what* is learnt (Mollon and Danilova, 1996). For instance, learning to discover information could come in the form of knowing what orientation channels to base decisions on and what orientation channels to ignore.

1.3 Selective processing during face perception

Upright face identification relies on horizontally oriented structures (Dakin and Watt, 2009; Goffaux and Dakin, 2010), and an observer's relative sensitivity to horizontal and vertical facial structure is referred to as "horizontal bias" (also known as horizon-

tal tuning/selectivity). Individual differences in horizontal bias are correlated with performance in face identification tasks (Pachai et al., 2013), and with the magnitude of the face inversion effect (Pachai et al., 2013), suggesting that our lifelong experience with upright faces manifests as a horizontal bias for upright but not inverted faces.

Individual differences in horizontal bias could be due to several factors, including age, clinical diagnoses, and experience. Indeed, horizontal bias changes across the lifespan (Goffaux et al., 2015; Obermeyer et al., 2012) and is weaker in people with prosopagnosia (Pachai et al., 2015). Horizontal bias is greater for familiar than unfamiliar faces (Pachai et al., 2017), reflecting more extensive perceptual learning for familiar than unfamiliar faces. Experience-based changes in horizontal bias may be used to reduce face perception deficits in different populations. For example, older adults have more difficulty identifying faces than younger adults (Grady et al., 1994; Konar et al., 2013), and they also show a decreased horizontal bias (Obermeyer et al., 2012). An important, unresolved issue is whether practice-related improvements in face identification are accomplished by increasing horizontal bias or by some other mechanism.

The observation of a horizontal bias in face perception provides a unique opportunity to study the neural substrates of orientation-selective processing in pre-defined neural markers of interest. Decades of research on face perception have established that faces elicit signature brain responses that often behave in predictable ways. For example, using electroencephalography (EEG) to measure the evoked response to a face has revealed the N170, a brain response larger to faces than non-face objects (Bentin et al., 1996; Rossion et al., 2000), and the N250, a brain response larger to familiar than unfamiliar faces (Tanaka et al., 2006; Kaufmann et al., 2009). Hence, these neural markers of face perception can be targeted when assessing the neural sensitivity to horizontal bias – if face processing relies on the strength of horizontal bias, then we can expect to see more face-typical brain responses to the horizontal structure of faces than to the vertical structure of faces (i.e., a neural horizontal bias).

1.4 Thesis Experiments

This dissertation investigates the role of information discovery in perceptual learning. We asked if information discovery can be encouraged through the use of orientation-selective processing, but first we establish the presence of orientation-selective process-

ing in a task that we already believe is done through orientation-selective processing. In Chapter 2, the orientation-selective processing thought to occur during face perception is measured behaviourally and with EEG, and we asked if there is a neural measure of horizontal bias that is correlated with face identification performance. Chapter 3 investigates how neural horizontal bias is affected by healthy ageing, which is a factor associated with face perception deficits. Chapter 4 uses perceptual learning to improve face identification accuracy in older adults, and assesses the effects of learning on horizontal bias. Next, Chapter 5 asks if perceptual learning of textures can be used to encourage the generalizable, orientation-selective processing of novel textures. Finally, Chapter 6 assesses how the learning found in Chapter 5 is affected by giving participants substantially more practice and therefore more time to discover the relevant structure. In summary, this dissertation asks two broad questions: 1) how is orientation-selective processing in faces reflected in neural correlates of perception; and 2) can observers learn to discover orientation structure in a novel stimulus class (i.e., textures). To anticipate the results, below are summaries of each chapter's findings.

In Chapter 2, we used EEG to record the neural representation of orientation-selective processing while perceiving faces, a stimulus class for which orientation-selective processing is known to occur. We confirmed that accurate face identification relies on the presence of horizontal facial structure more than vertical facial structure. We refer to the preferential use of horizontal relative to vertical structure as horizontal bias, and replicate a previous finding that horizontal bias is correlated with full face identification accuracy. We then show that neural signatures of face processing rely on the presence of horizontal facial structure. Critically, we isolate a neural marker of horizontal bias: the difference in evoked activity from processing horizontally versus vertically oriented structures, and found that the neural horizontal bias also was correlated with face identification abilities.

In Chapter 3, we investigated if deficits in face perception are reflected in behavioural and neural measures of horizontal bias. We collected behavioural and neural measures of horizontal bias from healthy older adults. Although face identification accuracy was lower in older than younger adults, accuracy was still significantly correlated with behavioural and neural measures of horizontal bias. Face identification accuracy in older adults was also correlated with neural horizontal bias at an earlier point in the evoked response than was found in younger adults. The results support the idea that orientation-selective processing is a critical part of accurate face identification, even

when face identification is affected by healthy ageing.

In Chapter 4, we investigated how older adult face identification changes with practice in a face identification task, and if those changes are associated with changes in the magnitude of horizontal bias. We measured the magnitude of horizontal bias before and after upright face identification training. We show that the performance improvement following the perceptual learning of upright faces was associated with an increased horizontal bias. Together, these results reinforce not only the importance of orientation-selective processing during face perception, but that older adults can re-learn to discover the informative structure in faces.

In Chapter 5, we investigated if observers could discover that stimuli contained informative structure in a select orientation band. Essentially, we asked if orientation-selective processing can emerge for a stimulus class that, unlike faces, are not naturally subject to orientation-selective processing. We manipulated structured noise patterns (i.e., textures) so that they included diagnostic information at only certain orientations, and non-diagnostic at others. For this stimulus set the optimal identification strategy would be to selectively process the diagnostic orientations. In six experiments, we show that observers can learn to make perceptual decisions which rely more on the diagnostic orientations. However, the amount of learning was small and was specific to the trained textures. These experiments point towards a potentially promising paradigm where observers learn to discover the task-relevant information of stimuli in which the informative and non-informative structures are not easily separable.

In Chapter 6, we extended the findings from Chapter 5 by providing observers with more opportunity to discover the informative structure in a texture identification task. We found that observers varied significantly in how much practice aided their response accuracy. Pre- to post-training assessment of their orientation-selectivity revealed that the fivefold increase in the amount of training promoted generalizable learning to novel textures, suggesting that observers were able to discover the informative structure of novel textures in which the learned strategy was applicable.

Taken together, this dissertation provides support for the idea that orientation-selective processing of faces is reflected in the visual system, and that observers can learn to discover structure in a stimulus according to the orientation of the structure. Specifically, we provide neural markers for orientation-selective processing when perceiving faces (Chapters 2 and 3), a stimulus class for which a lifelong of perceptual learning has resulted in a strong bias for horizontally oriented structures.

We also provide evidence that age-related deficits in orientation-selective processing of faces can be recovered (Chapter 4). We then move on to show that naive observers can learn to discover select orientation structures of textures according to its diagnostic value (Chapter 5), and that adequate practice applying the learned strategy can generalize to novel textures for which the orientation-selective processing is relevant (Chapter 6). These results inform our understanding of what information we have access to when making perceptual judgments, and how stimulus characteristics can dictate the strategy learned to successfully complete a perceptual task.

References

Ball, K. and Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6).

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, 8(6):551–565.

Birch, E. and Petrig, B. (1996). FPL and VEP measures of fusion, stereopsis and stereoacuity in normal infants. *Vision Research*, 36(9):1321–1327.

Dakin, S. C. and Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

Drew, T., Evans, K., Võ, M. L. H., Jacobson, F. L., and Wolfe, J. M. (2013). Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical Images? *RadioGraphics*, 33(1):263–274.

Fiorentini, A. and Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Research*, 21.

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

Goffaux, V. and Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, 1(143):1–14.

Goffaux, V., Poncin, A., and Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 Year Old). *PLoS ONE*, 10(9):1–17.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, 402:176–8.

Gold, J. M., Sekuler, A. B., and Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28:167–207.

Gosselin, F. and Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, 41(17):2261–2271.

Grady, C. L., Ma Maisog, J., Horwitz, B., Ungerleider, L. G., Mentis, M. J., Salerno, J. A., Pietrini, P., Wagner, E., Haxby, J. V., Gillette, J., Giacometti, K., Baldwin, P., Jacobs, G., Stein, S., Green, S., Fluck, S., and Der, M. (1994). Age-related Changes in Cortical Blood Flow Activation during Visual Processing of Faces and Location. *The Journal of Neuroscience*, 14(3):1450–1462.

Hadad, B., Schwartz, S., Maurer, D., and Lewis, T. L. (2015). Motion perception: a review of developmental changes and the role of early visual experience. *Frontiers in Integrative Neuroscience*, 9(September).

Hashemi, A., Lass, J. W., Truong, D., Sekuler, A. B., and Bennett, P. J. (2013). The time-course of rapid stimulus-specific perceptual learning. *Journal of Vision*, 13(9):1090.

Hussain, Z., Bennett, P. J., and Sekuler, A. B. (2012a). Versatile perceptual learning of textures after variable exposures. *Vision Research*, 61:89–94.

Hussain, Z., McGraw, P. V., Sekuler, A. B., and Bennett, P. J. (2012b). The rapid emergence of stimulus specific perceptual learning. *Frontiers in Psychology*, 3(July):226.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009a). Contrast-reversal abolishes perceptual learning. *Journal of Vision*, 9:1–8.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009b). How much practice is needed to produce perceptual learning? *Vision Research*, 49(21):2624–34.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2011). Superior identification of familiar visual patterns a year after learning. *Psychological Science*, 22(May):724–30.

Karni, A. and Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, 7:530–535.

Karni, A. and Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *PNAS*, 88(11):4966–70.

Karni, A. and Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(6443):250–2.

Kaufmann, J. M., Schweinberger, S. R., and Burton, A. M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience*, 21(4):625–41.

Konar, Y., Bennett, P. J., and Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision Research*, 88:38–46.

Lewis, T. L. and Maurer, D. (2005). Multiple sensitive periods in human visual development: Evidence from visually deprived children. *Developmental Psychobiology*, 46(3):163–183.

McKee, S. P. and Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception & Psychophysics*, 24(3):258–262.

Mollon, J. D. and Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision*, 10(1):51–58.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., and Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, 6(10):575–585.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., and Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125(8):1063–1068.

Obermeyer, S., Kolling, T., Schaich, A., and Knopf, M. (2012). Differences between old and young adults' ability to recognize human faces underlie processing of horizontal information. *Frontiers in Aging Neuroscience*, 4(April):3.

Pachai, M. V., Bennett, P. J., Sekuler, A. B., Corrow, S., and Barton, J. J. S. (2015). Sensitivity to horizontal structure and face identification in developmental prosopagnosia and healthy aging. *Canadian Journal of Experimental Psychology*, 69(4):341.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to information conveyed by horizontal contours is correlated with face identification accuracy. *Frontiers in Psychology*, 4(74).

Pachai, M. V., Sekuler, A. B., Bennett, P. J., Schyns, P. G., and Ramon, M. (2017). Personal familiarity enhances sensitivity to horizontal structure during processing of face identity. *Journal of Vision*, 17(6):5.

Peterson, M. and Eckstein, M. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109(48).

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., and Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11(1):69–74.

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research*, 51(13):1552–66.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–6.

Siu, C. R. and Murphy, K. M. (2018). The development of human visual cortex and clinical implications. *Eye and Brain*, 10:25–36.

Sowden, P. T., Davies, I. R. L., and Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):379–390.

Tanaka, J. W., Curran, T., Porterfield, A. L., and Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, 18(9):1488–97.

Chapter 2

The role of horizontal facial structure on the N170 and N250

2.1 Abstract

Recent studies have shown that horizontal facial structure is important for face identification. Also, sensitivity to horizontal structure is associated with the size of the face inversion effect. However, it is unclear how the N170 and N250, two components of visual event-related potentials ERPs that have been implicated in face perception, are modulated by oriented facial structure in an upright face identification task. Here, we recorded ERPs and behavioural accuracy from adult observers performing a 1-of-6 face identification task in conditions that parametrically manipulated the orientation structure of upright faces. Faces were filtered with ideal orientation filters centred on either 0 (horizontal) or 90 deg (vertical). Filter bandwidth was varied across conditions from ±45 to ±90 deg in steps of ±9 deg. As has been reported previously, response accuracy was significantly higher for faces that contained horizontal structure than vertical structure, and the horizontal-vertical difference was correlated with accuracy for unfiltered faces. In addition, the N170 and N250 were affected by the manipulation of horizontal facial structure. Furthermore, for the N250, but not the N170, the relative sensitivity to horizontal compared to vertical facial structure was significantly correlated with identification accuracy for unfiltered faces. We suggest that in a face identification task, the N250 but not the N170 is modulated by the amount of diagnostic information conveyed by horizontal structure.

2.2 Introduction

Human adults can identify faces that differ in viewpoint, illumination, and expression, but have great difficulty identifying faces that are rotated 180 deg in the picture plane (Yin, 1969). This so-called face inversion effect (FIE) likely is the by-product of everyday experience with upright, but not inverted faces (Valentine, 1988; Sekuler et al., 2004), and the influence of experience on recognition has been demonstrated in several studies that have shown that practice can induce large inversion effects for non-face objects (Diamond and Carey, 1986; Gauthier et al., 2000; Gauthier and Bukach, 2007; Husk et al., 2007; Hussain et al., 2009).

Why does inversion impair face identification? One idea is that experience with upright faces leads to the development, or improvement, of holistic/configural processing which complements feature-based processing, but which is disrupted by stimulus inversion (Tanaka and Farah, 1993; Farah et al., 1998; Maurer et al., 2002). However, the configural/feature framework has been hindered by a lack of consensus about what constitutes a feature or a configural cue (Piepers and Robbins, 2012). In addition, Konar et al. (2010) showed that the inversion effect could not be accounted for by configural processing. Another idea is that qualitatively similar strategies are used to encode upright and inverted faces (Sekuler et al., 2004, Willenbockel et al., 2010, Murphy and Cook, 2017), but that experience and learning increases efficiency for upright faces relative to inverted faces (Gold et al., 1999a, 2004). According to this hypothesis, if information at particular spatial scales was critical for face identification, then we might expect to see differences in the spatial frequency selectivity for upright and inverted face identification. If observers used different spatial scales to identify upright vs. inverted faces, one would expect to find differences in the spatial frequency tuning for upright and inverted faces. In fact, human observers do rely most heavily on facial information in the 8-13 cyc/face spatial frequency range when discriminating upright faces (Näsänen, 1999; Gold et al., 1999b; Tanskanen et al., 2005; Gaspar et al., 2008; Keil et al., 2008); however, Gaspar et al. (2008) found no difference in spatial frequency selectivity for upright and inverted faces (also see Willenbockel et al., 2010; Royer et al., 2017). Another possibility is that observers use different spatial sampling strategies to collect information about upright and inverted faces; however, studies using classification images (Sekuler et al., 2004) and eye-tracking (Williams and Henderson, 2007; Rodger et al., 2009) have found that observers identify both

upright and inverted faces using information primarily conveyed by pixels around the eyes and brows. Hence, previous studies have failed to find dramatic effects of face orientation on spatial frequency selectivity or spatial sampling of facial information.

Researchers recently have begun studying how observers use information conveyed by oriented facial structure. Dakin and Watt (2009) noted that the most diagnostic information of a face's identity is carried by structure in a horizontal orientation band centered around 0 deg (Figure 2.1a). Goffaux and Dakin (2010) demonstrated that selectively removing horizontal facial structure, but not vertical structure, significantly affected the face inversion effect, facial identity after-effect, face-matching across different viewpoints, and certain measures of holistic processing. Furthermore, Pachai et al. (2013) found that greater sensitivity to horizontal relative to vertical structure was significantly correlated with overall face identification accuracy for upright but not inverted faces, and that the degree of horizontal tuning was correlated with the magnitude of the face inversion effect. These results suggest that differential sensitivity to horizontal facial structure may underlie the face inversion effect. More recently, sensitivity to horizontal facial structure has been shown to contribute to more accurate identification of familiar compared to unfamiliar faces (Pachai et al., 2017), whereas reduced sensitivity to horizontal structure has been linked to prosopagnosia (Pachai et al., 2015). Sensitivity to horizontal facial structure also plays an important role in the perception of emotional expressions (Huynh and Balas, 2014; Balas et al., 2015a), and changes in horizontal tuning have been linked to changes in face perception that occur during childhood (Goffaux et al., 2015; Balas et al., 2015b), normal ageing (Obermeyer et al., 2012; Sekuler et al., 2015; de Heering et al., 2016), and central vision loss (Yu and Chung, 2011).

Horizontal tuning, or bias, also has been found in neural mechanisms underlying face processing. For example, preferential activation to horizontal compared to vertical facial structure has been found in neurons in the lateral anterior patch of monkey inferior temporal cortex (Taubert et al., 2016), and in BOLD activation in the human fusiform face area (Goffaux et al., 2016). Jacques et al. (2014) examined the influence of horizontal structure on the N170, the earliest time-window in the event-related potential (ERP) that is differentially sensitive to faces (Bentin et al., 1996; Rossion et al., 2000; Rousselet et al., 2004). The N170, like behavioural measures, exhibits a face inversion effect. That is to say, there is a reliable difference between N170 latency and/or amplitude evoked by upright and inverted faces (Bentin et al., 1996; Rossion

et al., 2000; Eimer, 2000; de Haan et al., 2002; Itier and Taylor, 2002; Rousselet et al., 2004, 2008a). Interestingly, the N170 and behavioural FIEs are correlated (Jacques and Rossion, 2007), which suggests that similar changes in perceptual processing may underlie the behavioural and ERP measures. Consistent with this idea, Jacques et al. (2014) found that the N170 inversion effect, like the behavioural effect, depends on horizontal facial structure remaining intact. Specifically, they found that phase-scrambling horizontal structure, but not vertical structure, in upright faces yielded N170s that differed from N170s evoked by normal, unscrambled faces.

Using the inversion effect as an index of intact face perception, Jacques et al. (2014) determined that horizontal structure is required for the N170 latency inversion effect, much as it is required for the behavioural face inversion effect (Dakin and Watt, 2009; Goffaux and Dakin, 2010; Pachai et al., 2013, 2018). However, the exact contribution of horizontal structure in eliciting an N170 during upright face *identification* remains unclear. How does the N170 to an upright face change as a function of the orientation structure present? How are horizontal biases in electrophysiological and behavioural measures related? The current study examined these questions by measuring the effects of orientation filtering on face identification accuracy and the amplitude and latency of the N170.

The N170 is the earliest ERP for faces but, its relation to face identification is unclear: Although the N170 may habituate to repeated presentations of a single face identity (Heisz et al., 2006), it generally is not sensitive to face identity (Eimer, 2000; Amihai et al., 2011). A slightly later ERP component, the N250, is more strongly associated with face identity (Schweinberger et al., 2002, Schweinberger et al., 2004). Tanaka and colleagues suggested that the N250 is the earliest memory-related component for object individuation (Scott et al., 2008), including face recognition (Tanaka et al., 2006). After repeated exposures, the N250 is largest when stimuli are upright faces, compared to inverted faces or cars (Schweinberger et al., 2004). Additionally, the N250 is enhanced for repetitions of the same face identity but different viewpoints (Kaufmann et al., 2009), implicating the N250 in the overall acquisition of facial identity. Tanaka et al. (2006) further showed that the N250 both was strongest for long-term familiar faces and for newly learned identities. Based on these results, we hypothesized that the N250 should be strongest when the facial identity is clearly observed (i.e., when horizontal structure is most preserved), and less pronounced when the identity is not clear (i.e., when horizontal structure is removed).

We compared ERPs elicited by unfiltered upright faces to ERPs evoked by orientation-filtered faces with varying degrees of horizontal and vertical structure. By parametrically increasing the amount of facial structure at each orientation, we hoped to understand what constitutes a typical N170 and N250.

2.3 Methods

2.3.1 Subjects

Twelve students from McMaster University participated in the experiment. One female participant's EEG data were excessively noisy, leaving eleven subjects for the analysis (6 males; range = 18-30 years old, M = 22.4, SD = 3.88). All were right-handed and had normal or corrected-to-normal Snellen acuity. Informed consent was obtained from each participant. Participants were reimbursed $10/hour or given partial course credit for participating. The experimental protocol was approved by the McMaster University Research Ethics Board.

2.3.2 Apparatus & Stimuli

Stimuli were generated on an Apple Macintosh G5 computer using MATLAB (Mathworks, 2007) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Stimuli were presented on a 21-inch ViewSonic G225f display with a resolution of 1280 × 1024 pixels (32 pixels/cm) and a 85 Hz refresh rate. Viewing was binocular through natural pupils from a viewing distance of 100 cm. The stimuli were centered on a 256 × 256 pixel matrix and subtended 4.6° of visual angle. The stimuli were constructed from a set of 6 faces (3 male), all having the same amplitude spectrum, selected from the set of 10 faces used by Gold et al. (1999b). Each face was filtered with an ideal band-pass orientation filter centered on either 0° (horizontal) or 90° (vertical). Orientation bandwidth varied from ±45 deg to ±90 deg in steps of ±9 deg, which resulted in a total of 11 conditions: five conditions using filters centered on 0 deg, five using filters centered on 90 deg, and one using an unfiltered face. Face RMS contrast was 0.05 prior to filtering. After filtering, RMS contrast varied with filter bandwidth (Table 2.1). Note that at all bandwidths, RMS contrast was ≈ 10% higher for stimuli constructed with vertical filters than horizontal filters.

Bandwidth (deg)	Vertical Filter	Horizontal Filter
±45	0.0372	0.0334
±54	0.0400	0.0359
±63	0.0408	0.0376
±72	0.0442	0.0407
±81	0.0455	0.0418
±90	0.0500	0.0500

Table 2.1 – RMS contrast for for stimuli constructed with orientation filters of various bandwidths centered on 90 (vertical) or 0 deg (horizontal). For each filter, all six face identities had the same RMS contrast. The ±90 bandwidth passed all orientations, and therefore faces in that condition were identical to unfiltered faces.

2.3.3 Procedure

The task was a six alternative forced choice (6-AFC) identification task. The experiment began with 10 practice trials using only unfiltered faces to avoid prior exposure to filtered faces before the experimental trials. Each trial started with the presentation of a fixation point in the center of the display. After a random duration that varied from 0.9 to 1.1 s, the fixation point was extinguished and immediately followed by the presentation of a face stimulus for 0.2 s. The stimulus was replaced by a blank uniform screen for 1 s, and then by a response screen that contained all six faces (Figure 2.1b). The response screen images were always unfiltered and had an RMS contrast of 0.3. The participant indicated his/her response by clicking on one of the faces with a computer mouse. Auditory feedback was provided after every response in the form of 600 and 200 Hz tones after correct and incorrect responses, respectively. The next trial started 0.25 s following the response. We did not record response times.

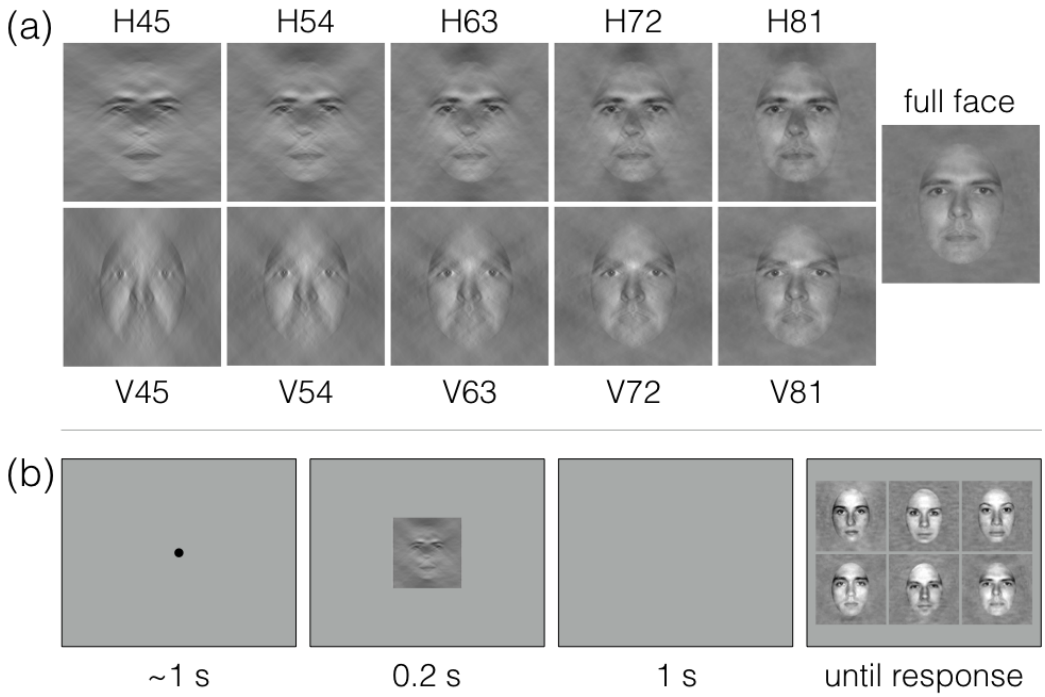

Figure 2.1 – **A)** Examples of stimuli constructed from a single identity in all filter conditions. Horizontal (H) and vertical (V) filters had bandwidths ranging from ±45° to ±81°. The far right image is the full (i.e., unfiltered) face. **B)** Sample trial structure. On each trial, a random temporal jitter between -100 and 100 ms was added to the fixation point duration.

2.3.4 Design

In each 40 minute session, participants were presented with each face seven times in 11 conditions, yielding a total of $6 \times 7 \times 11 = 462$ experimental trials per session. Each participant completed two sessions, or 924 total trials (i.e., 84 trials per condition). All conditions and faces were randomized across trials. Each session contained three brief rest periods. There was no *a priori* hypothesis about sessions, and to improve the signal-to-noise ratio of the ERPs, data were collapsed across the two days prior to conducting statistical analyses.

2.3.5 Electrophysiology

EEG data were acquired during the behavioural task using a 256-channel HydroCel Geodesic Sensor Net (Electric Geodesics Inc., Eugene, Oregon; Tucker, 1993). Data were referenced online to electrode Cz, and sampled at 500 Hz. Offline, data from each trial were segmented from -200 ms to 998 ms and transferred to MATLAB for further processing using the EEGLAB (Delorme and Makeig, 2004) and *LIMO EEG* (Pernet et al., 2011) toolboxes. The responses to faces and objects occur primarily in the 5-15 Hz range (Rousselet et al., 2007a), and therefore we applied a 30 Hz low-pass non-causal filter to the EEG data using *pop_eegfiltnew*. No high-pass filter was used. Baseline correction was applied using the average of the 200 ms of pre-stimulus activity. Trials with amplitudes above +100 μV or below -100 μV were rejected. No other artifact rejection was applied. Unless otherwise noted, statistical analysis was restricted to the average of two clusters of 15 electrodes, one cluster from each hemisphere, centred on electrodes PO7 and PO8. The average of clusters was used instead of a single electrodes to accommodate for the variable fit of the geodesic nets – that is, the same electrode was not necessarily at the same anatomical spot across participants, but was always within the cluster we chose to average across. Nonetheless, the pattern of results were unchanged if we only used electrodes PO7 and PO8.

2.3.6 EEG Analysis

N170 mean amplitude was measured as the average voltage in a 42 ms time-window centred on the N170 peak of the grand-average ERP for each condition and hemisphere. By centering the time-window on each condition's grand-average latency, we reduced

the influence of latency differences across conditions on our measure of amplitude (Luck, 2005). The P100 mean amplitude was measured using the same technique, but a smaller time-window of 22 ms centered on the P100 peak identified in the grand-average ERP for each condition and hemisphere.

Latency was measured as the time of the N170 peak, which was automatically selected as the lowest local peak found in the 140-220 ms time-window. The P100 peak latency was measured using the same technique, but between 80-140 ms, and was defined as the highest local peak. All peak selections were verified manually.

It was much more difficult to identify the peak of the N250, so a single time window of 250-300 ms was used to quantify the N250 mean amplitude for all conditions (Schweinberger et al., 2004, Kaufmann et al., 2009). Given the difficulty in identifying a discrete peak, no measure of N250 latency was computed.

2.4 Results

2.4.1 Behaviour

To maintain an orthogonal design, the full face condition was omitted from all analyses of variance (ANOVAs), unless otherwise noted. Where appropriate, the Bonferroni correction for multiple comparisons was used to control familywise Type I error rate. For any main effects and interactions including *bandwidth* as a factor, we did not assume sphericity and report Huynh-Feldt corrected p-values instead. Effect size was measured using generalized eta squared (η^2_G), as described by Bakeman (2005) and Olejnik and Algina (2003) for repeated-measure designs. Lakens (2013) suggests that Cohen (1988)'s suggested benchmarks of effect size can be used when interpreting η^2_G (small: 0.01, medium: 0.04, large: 0.16). For t tests, we report effect size using Cohen's d. Statistical analyses were performed with R (R Core Team, 2017).

Response accuracy is plotted for each condition in Figure 2.2, which shows that response accuracy generally increased as filter bandwidth increased from ±45 to ±81, and that the increase was greater for vertical filters (V) than horizontal filters (H). Proportion correct was submitted to a repeated-measures ANOVA with factors of *filter orientation* (horizontal and vertical) and *bandwidth* (±45, 54, 63, 72, and 81 degrees). The main effects of *filter orientation* ($F_{(1,10)} = 180$, $p < 0.0001$, $\eta^2_G = 0.63$) and *bandwidth* ($F_{(4,40)} = 189$, $p_{HF} < 0.0001$, $\eta^2_G = 0.50$) were significant, as was the

filter orientation × *bandwidth* interaction ($F_{(4,40)} = 37.2$, $p_{HF} < 0.0001$, $\eta^2_G = 0.25$). The interaction was significant because the effect of bandwidth was much larger with vertical filters ($F_{(4,40)} = 159$, $p < 0.0001$, $\eta^2_G = 0.718$) than with horizontal filters ($F_{(4,40)} = 14.9$, $p < 0.0001$, $\eta^2_G = 0.162$). For instance, response accuracy improved by 48% from the smallest (V45) to the largest (V81) vertical filter bandwidth, but improved by only 14% from the smallest (H45) to the largest (H81) horizontal filter. Note that in the V81 condition, the bandwidth was so large that all orientations except those within ±9 deg of horizontal were passed by the "vertical" filter. Thus, the conditions yielding the highest response accuracy always had considerable horizontal facial structure. Consistent with this idea, accuracy with the narrowest horizontal filter (H45, $M = 0.74$) was higher than accuracy in the V72 ($M = 0.64$) condition ($t_{10} = 3.17$, $p = 0.0101$, $d = 0.955$) and did not differ significantly from accuracy in the V81 ($M = 0.81$) condition ($t_{10} = -2.04$, $p = 0.0686$, $d = 0.615$).

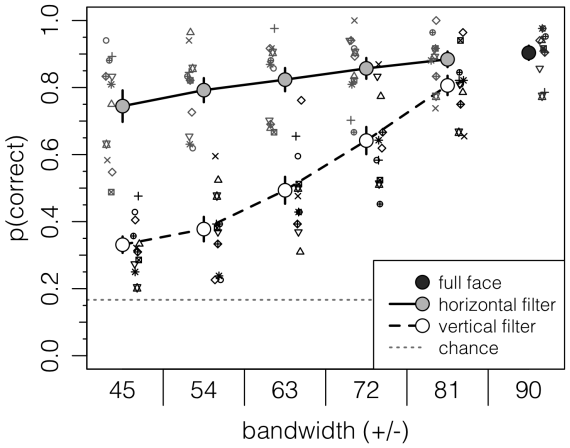

Figure 2.2 – Mean proportion correct plotted as a function of *bandwidth* (degrees), separately for the horizontal (grey) and vertical (white) *filter orientations*, and the full face condition (black). Individual subject results are also plotted, jittered for legibility. Error bars are ± 1 SEM. Chance performance in the 6-AFC task is indicated with a horizontal dotted line. The effect of *bandwidth* is noticeably larger for the vertical filter orientation than the horizontal filter orientation, and accuracy in all conditions except the two that used horizontal filters with bandwidths of 72 and 81 deg was significantly different than from accuracy in the full face condition.

Unfiltered faces contain information at all orientations, so we expected accuracy to be highest in that condition. Inspection of Figure 2.2 is consistent with that expectation: accuracy in most conditions that used horizontally-filtered faces, and all conditions that used vertically-filtered faces, was lower than accuracy obtained with

unfiltered faces. To evaluate this idea, we used one-tailed paired t tests to compare accuracy in each filtered-face condition to accuracy in the unfiltered face condition. We used the Bonferroni method to maintain a familywise Type I error rate of 0.05 by setting the per-comparison α to 0.005. Compared to full face identification accuracy ($M = 0.90$), accuracy with horizontally-filtered faces was significantly lower in the H45 ($M = 0.74$, $t_{10} = -4.85$, $p = 0.00034$, $d = 1.46$), H54 ($M = 0.79$, $t_{10} = -5.50$, $p = 0.00013$, $d = 1.66$), and H63 ($M = 0.82$, $t_{10} = -3.69$, $p = 0.0021$, $d = 1.11$) conditions, but not in the H72 ($M = 0.86$, $t_{10} = -3.14$, $p = 0.0053$, $d = 0.95$) or H81 ($M = 0.88$, $t_{10} = -1.66$, $p = 0.0637$, $d = 0.50$) conditions. In contrast, accuracy in all of the vertically-filtered conditions was significantly less than accuracy in the full face condition (V45: $M = 0.33$, $t_{10} = -31.1$, $p < 0.0001$, $d = 9.38$; V54: $M = 0.38$, $t_{10} = -25.4$, $p < 0.0001$, $d = 7.65$; V63: $M = 0.49$, $t_{10} = -15,1$, $p < 0.0001$, $d = 4.56$; V72: $M = 0.64$, $t_{10} = -9.32$, $p < 0.0001$, $d = 2.81$; and V81: $M = 0.81$, $t_{10} = -6.51$, $p < 0.0001$, $d = 1.96$). Differences between filtered and unfiltered faces were noticeably larger for the vertically-filtered than horizontally-filtered conditions.

2.4.2 Event-Related Potentials

Spatiotemporal Analyses

We started our EEG analysis by inspecting the differences between the spatiotemporal responses obtained in three main conditions. First, we subtracted the mean ERP in the full face condition from the mean ERP obtained in the H45 condition. This comparison, illustrated in Figure 2.3a, reveals differential activation when only horizontal structure is present compared to when all orientation structure is present. We did the same comparison between ERPs obtained in the V45 and full face conditions (Figure 2.3b) to reveal differential activation when only vertical structure is present compared to when all orientation structure is present. Finally, we computed the difference between ERPs in the V45 and H45 conditions (Figure 2.3c). This third comparison is between two conditions that used stimuli that contained structure that fell within a 90 deg band, and therefore highlights differences between ERPs evoked by vertical and horizontal facial structure without possible confounds due to differences in stimulus bandwidth.

The topographic difference plots in Figure 2.3a and 2.3b reveal that compared to H45 and V45, the full faces produced a significantly more negative ERP in the N170 (≈150-200 ms) and N250 (≈250-300 ms) time windows. These differences were

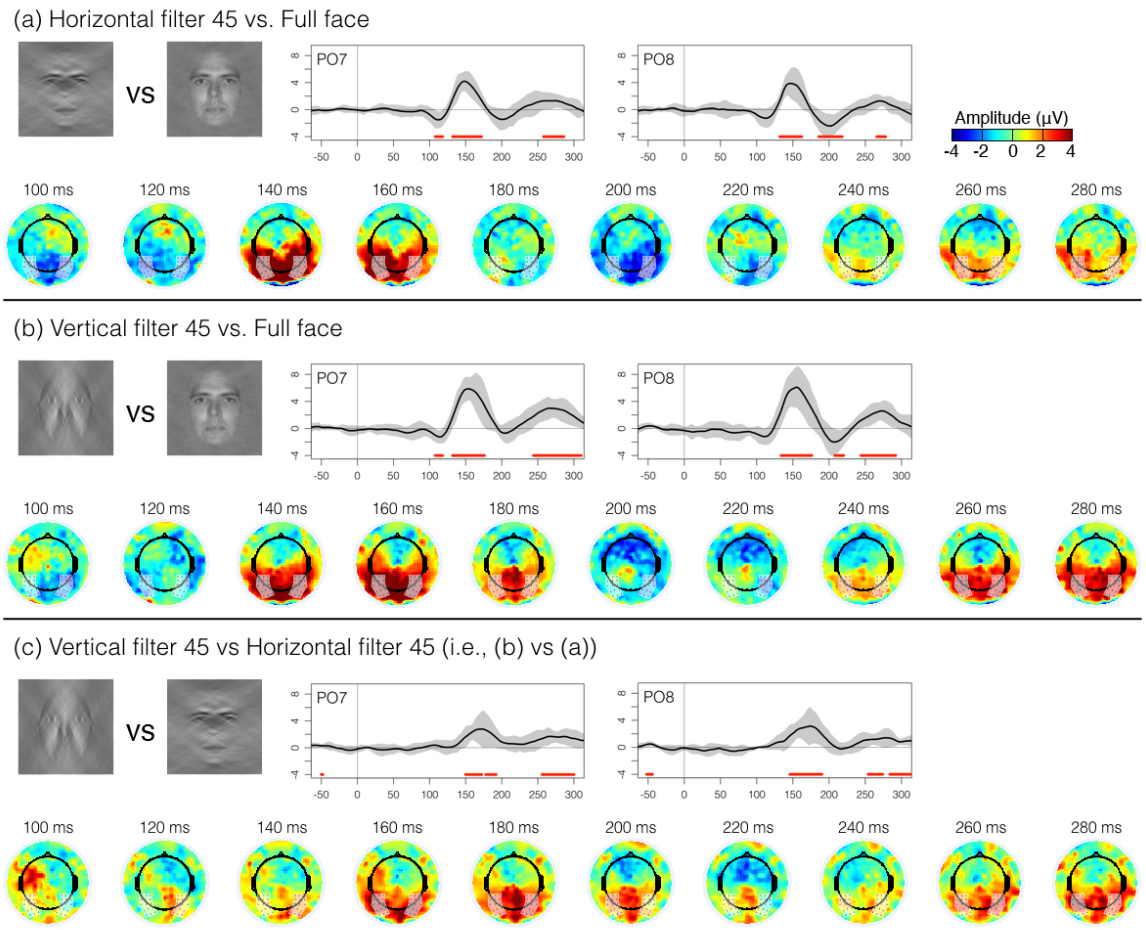

Figure 2.3 – Summary of ERP *differences* between the **A)** H45 and full face conditions; **B)** V45 and full face conditions; and **C)** H45 and V45 conditions. Each panel shows the conditions being compared (left image minus right image), and the grand-average difference waves with 95% confidence intervals calculated for electrodes PO7 and PO8. Difference waves were evaluated at each point with t tests; significant ($p < 0.05$, uncorrected) deviations from zero are indicated by red dots on the bottom of each difference-wave plot. The bottom of each panel contains a series of topographic plots from 100 ms to 280 ms after stimulus-onset, in 20 ms intervals. The colour scale for the topographic plots is shown on the right in (a). White shaded regions in the topographic plots represent the two 15-electrode clusters averaged for analysis. The LIMO EEG toolbox (Pernet et al., 2011) was used to construct topographic plots, estimate confidence intervals, and perform the t tests.

broadly distributed in the posterior/occipital region. To visualize the effect in a more typical ERP, we plotted the ERP differences at the PO7 and PO8 electrodes for each comparison. The comparison between H45 and V45 (Figure 2.3c) revealed that the two filtered conditions also differed significantly in the N170 and N250 time windows. These differences were, again, broadly distributed in the posterior/occipital region.

The analyses illustrated in Figure 2.3 were done to visualize the spatiotemporal distribution of the differential activity seen across critical conditions. As such, no corrections were done, nor were any additional analyses of the entire time-series across the entire topography. We focused the remainder of our analysis on the N170 and N250 time windows, and on two clusters of electrodes centred on PO7 and PO8. As a control, we also analyzed the P100 time window at the same regions.

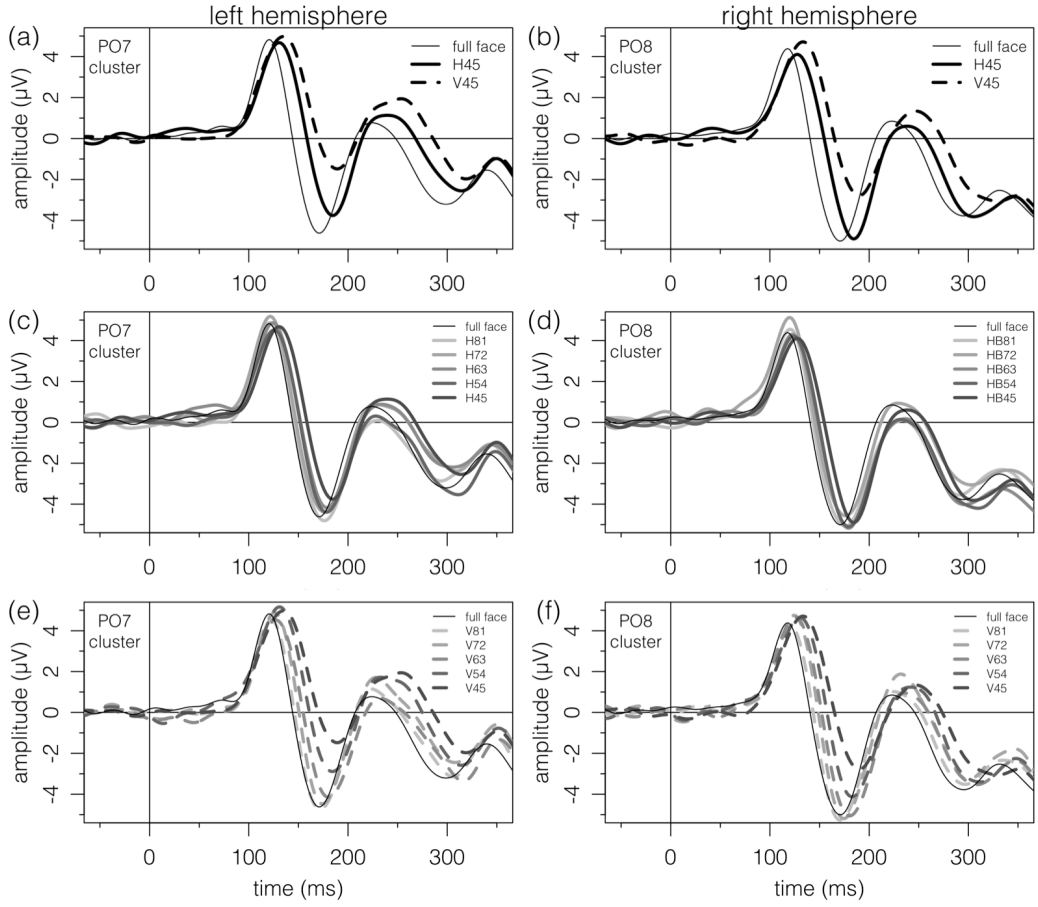

Figure 2.4 – Grand average ERP traces plotted separately for the left and right hemispheres. **A,B)** Full face (grey), Horizontal Base 45 (black solid), and Vertical Base 45 (dashed) are shown in a single plot to reveal how affected the N170 is when horizontal structure is omitted (VB45) compared to when vertical structure is omitted (HB45). There is a noticeable difference between parametrically adding vertical structure to a horizontal base **(C,D)** and adding horizontal structure to a vertical base **(E,F)**, demonstrating how it is the presence of horizontal structure that primarily modulates the N170. The presence or absence of vertical structure affects the N170 minimally.

N170 amplitude & latency

Average ERPs recorded at PO7- and PO8-centered clusters in the left and right hemispheres are plotted in Figure 2.4. Inspection of Figure 2.4 suggests that N170 amplitude decreased when horizontal structure was removed from faces, and that amplitude was affected more by the manipulation of horizontal structure than vertical structure. These trends are more evident in Figure 2.5 a&b, which plot mean N170 amplitudes measured in the left and right hemispheres as a function of filter orientation and bandwidth. N170 amplitudes were analyzed with a 2 (*filter orientation*) × 5 (*bandwidth*) × 2 (*hemisphere*) ANOVA. The main effects of *filter orientation* ($F_{(1,10)} = 18.9$, $p = 0.0015$, $\eta^2_G = 0.007$) and *bandwidth* ($F_{(4,40)} = 7.30$, $p_{HF} = 0.0015$, $\eta^2_G = 0.022$) were significant. These main effects were qualified by a significant *orientation* × *bandwidth* interaction ($F_{(4,40)} = 7.39$, $p_{HF} = 0.0007$, $\eta^2_G = 0.010$). The interaction was significant because the effect of bandwidth was much larger with vertical filters ($F_{(4,40)} = 12.3$, $p_{HF} < 0.0001$, $\eta^2_G = 0.064$) than with horizontal filters ($F_{(4,40)} = 1.15$, $p_{HF} = 0.34$, $\eta^2_G = 0.0005$). There was no main effect of *hemisphere* ($F_{(1,10)} = 0.99$, $p = 0.344$, $\eta^2_G = 0.009$), nor did it interact with *filter orientation* ($F_{(1,10)} = 0.576$, $p = 0.466$, $\eta^2_G = 0.0002$). There was a significant *hemisphere* × *bandwidth* interaction ($F_{(4,40)} = 5.23$, $p_{HF} = 0.0050$, $\eta^2_G = 0.0013$), driven by a stronger bandwidth effect in the left ($F_{(4,40)} = 8.52$, $p_{HF} = 0.00016$, $\eta^2_G = 0.033$) than right hemisphere ($F_{(4,40)} = 5.29$, $p_{HF} = 0.0016$, $\eta^2_G = 0.015$). The three way interaction was not significant ($F_{(4,40)} = 0.561$, $p_{HF} = 0.688$, $\eta^2_G = 0.0002$). Finally, we used paired t tests to test if the two conditions with the narrowest filters produced N170 amplitudes that differed from the amplitude produced by the unfiltered face, or each other ($\alpha_{corrected} = 0.05/6 = 0.0083$). In the left hemisphere, the N170 to unfiltered faces ($M = -3.67\,\mu V$) did not differ from the N170 to H45 stimuli ($M = -2.89$, $t_{10} = 1.81$, $p = 0.100$, $d = 0.546$), but was significantly larger than the N170 to V45 stimuli ($M = -0.914$, $t_{10} = 6.70$, $p < 0.0001$, $d = 2.02$), and the N170 amplitude was larger for H45 than V45 stimuli ($t_{10} = 5.77$, $p = 0.00018$, $d = 1.44$). The results were similar in the right hemisphere: N170 amplitude to unfiltered faces ($M = -4.13$) was not significantly different than H45 ($M = -3.91$, $t_{10} = 0.532$, $p = 0.606$, $d = 0.16$), but it was significantly more negative than V45 ($M = -2.11$, $t_{10} = 5.51$, $p = 0.00026$, $d = 1.66$), and the N170 amplitude was larger for H45 than V45 stimuli ($t_{10} = 4.39$, $p = 0.0013$, $d = 1.32$).

N170 peak latencies are presented in Figure 2.5 c&d. In some respects the latency

results were similar to those obtained with N170 amplitude. In particular, N170 latency was affected by filter bandwidth and the effect appeared stronger for vertical compared to horizontal filters. The latency data were analyzed with an ANOVA that was identical to the one used to analyze N170 amplitudes. The main effect of *filter orientation* on peak latency was not significant ($F_{(1,10)} = 1.32$, $p = 0.28$, $\eta^2_G = 0.005$); however, the ANOVA did reveal a significant main effect of *bandwidth* ($F_{(4,40)} = 24.4$, $p_{HF} < 0.0001$, $\eta^2_G = 0.15$) and a significant *filter orientation* × *bandwidth* interaction ($F_{(4,40)} = 8.66$, $p_{HF} = 0.0003$, $\eta^2_G = 0.036$). The interaction was significant because the effect of *bandwidth* was greater for vertical filters ($F_{(4,40)} = 33.9$, $p_{HF} < 0.0001$, $\eta^2_G = 0.266$) than horizontal filters ($F_{(4,40)} = 5.16$, $p_{HF} = 0.0036$, $\eta^2_G = 0.072$). There was no significant main effect of *hemisphere* ($F_{(1,10)} = 0.931$, $p = 0.357$, $\eta^2_G = 0.0056$), nor did it interact with *filter orientation* ($F_{(1,10)} = 3.89$, $p = 0.077$, $\eta^2_G = 0.0015$) or *bandwidth* ($F_{(4,40)} = 2.42$, $p_{HF} = 0.110$, $\eta^2_G = 0.0057$). The three-way interaction also was not significant ($F_{(4,40)} = 1.66$, $p_{HF} = 0.179$, $\eta^2_G = 0.0015$). Finally, we used paired t tests ($\alpha_{corrected} = 0.0083$) to test if the two conditions with the narrowest filters produced N170 latencies that differed from each other and from the latency for unfiltered faces. In the left hemisphere, N170 latency was shorter for unfiltered faces ($M = 172.5$ ms) than for H45 ($M = 182.9$, $t_{10} = 3.99$, $p = 0.00256$, $d = 1.20$) and V45 stimuli ($M = 188.4$, $t_{10} = 4.63$, $p = 0.00094$, $d = 1.40$). N170 latencies for H45 and V45 stimuli did not differ significantly ($t_{10} = 2.29$, $p = 0.0451$, $d = 0.69$). The results were similar in the right hemisphere: N170 latency was significantly shorter for unfiltered faces ($M = 168.4$) than H45 ($M = 182.0$, $t_{10} = 9.11$, $p < 0.0001$, $d = 2.75$) and V45 stimuli ($M = 188.5$, $t_{10} = 7.96$, $p < 0.0001$, $d = 2.40$), and N170 latency was significantly shorter for H45 than V45 stimuli ($t_{10} = 3.43$, $p = 0.0064$, $d = 1.03$).

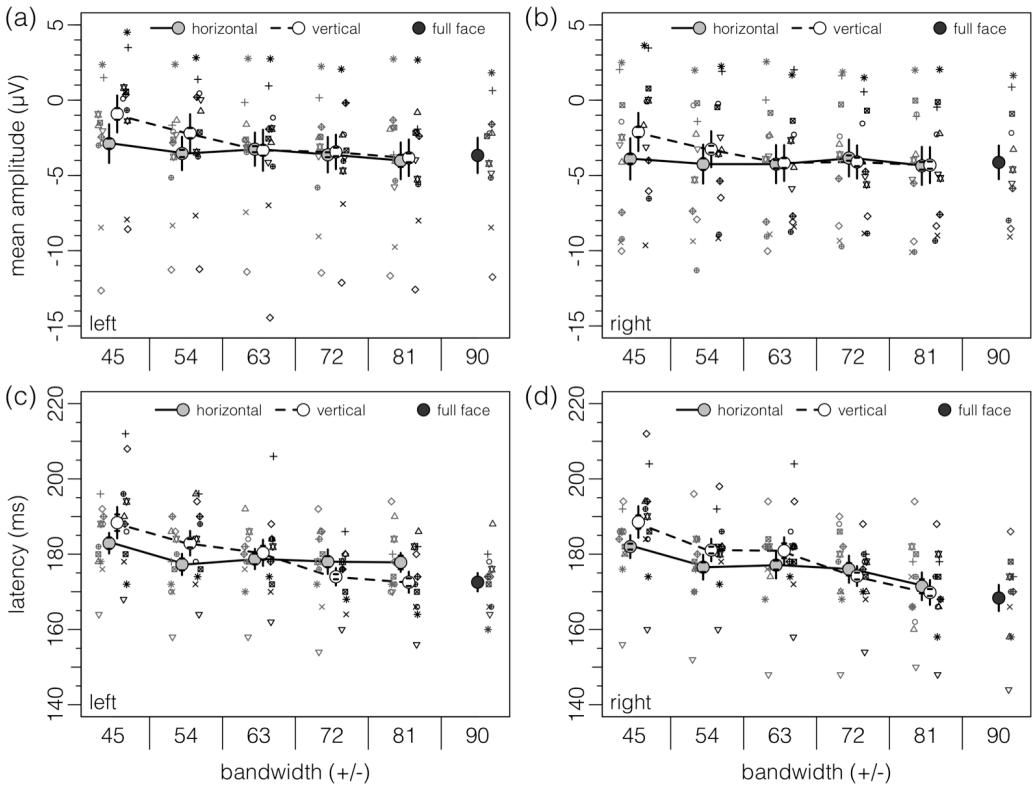

Figure 2.5 – N170 mean amplitude (top) and latency (bottom) averaged across all subjects, separately for left and right hemispheres. Error bars represent ± 1 SEM. Due to the inherently large between-subject variance in ERP data, and the repeated-measure design of the experiment, within-subjects corrected SEMs are also plotted as horizontal ticks on the SEM bars, as described by Loftus and Masson (1994), Cousineau (2005), and Morey (2008). Individual subject data also is plotted, and is jittered for legibility. Symbols are unique to each observer and preserved across conditions and figures.

N250 amplitude

N250 mean amplitudes are presented in Figure 2.6. N250 mean amplitude was more negative for the horizontal filter than vertical filter conditions. The amplitude became more negative as the vertical filter's bandwidth was increased, but was approximately constant as the horizontal filter's bandwidth increased. This pattern of results resembles the effects of stimulus filtering on N170 amplitude. N250 amplitudes were analyzed with a 2 (*filter orientation*) × 5 (*bandwidth*) × 2 (*hemisphere*) ANOVA. As with N170 amplitude, there were significant main effects of *filter orientation* ($F_{(1,10)} = 11.3$, $p = 0.007$, $\eta^2_G = 0.015$) and *bandwidth* ($F_{(4,40)} = 5.06$, $p_{HF} = 0.0026$, $\eta^2_G = 0.023$). The *filter orientation* × *bandwidth* interaction also was significant ($F_{(4,40)} = 4.14$, $p_{HF} = 0.0075$, $\eta^2_G = 0.007$), reflecting the fact that the effect of *bandwidth* was larger for vertical filters ($F_{(4,40)} = 6.59$, $p_{HF} < 0.0008$, $\eta^2_G = 0.053$) than horizontal filters ($F_{(4,40)} = 2.93$, $p_{HF} = 0.037$, $\eta^2_G = 0.018$). There was no significant main effect of *hemisphere* ($F_{(1,10)} = 1.60$, $p = 0.235$, $\eta^2_G = 0.0165$), nor did *hemisphere* interact significantly with *filter orientation* ($F_{(1,10)} = 0.222$, $p = 0.65$, $\eta^2_G = 0.0001$) or *bandwidth* ($F_{(4,40)} = 1.92, p_{HF} = 0.130, \eta^2_G = 0.0009$). The three way interaction also was not significant ($F_{(4,40)} = 1.94, p_{HF} = 0.151, \eta^2_G = 0.0016$). Finally, we used paired t tests ($\alpha_{corrected} = 0.0083$) to test if the conditions using the two narrowest filters produced N250 amplitudes different than the unfiltered face, or from each other. In the left hemisphere, the N250 amplitude evoked by unfiltered faces ($M = -2.04\,\mu V$) was significantly larger (more negative) than the N250 evoked by V45 stimuli ($M = 0.727$, $t_{10} = 5.26$, $p = 0.00037$, $d = 1.56$), but not H45 stimuli ($M = -0.57$, $t_{10} = 2.99$, $p = 0.0135$, $d = 0.902$). Also, the N250 was significantly more negative for H45 than V45 stimuli ($t_{10} = 3.33$, $p = 0.0076$, $d = 1.01$). The results were similar in the right hemisphere: N250 amplitude to unfiltered faces ($M = -2.62$) was significantly more negative than the N250 amplitude to V45 stimuli ($M = -0.319$, $t_{10} = 6.06$, $p = 0.00012$, $d = 1.83$), but did not differ significantly from the N250 amplitude to H45 stimuli ($M = -1.76$, $t_{10} = 2.78$, $p = 0.0193$, $d = 0.840$).

A reviewer suggested that our measure of the N250 1) was potentially too early and therefore may have included the P200 component; and 2) should be compared across sessions/blocks to capture effects of learning/familiarity, since the N250 is known to be sensitive to face familiarity (Kaufmann et al., 2009). To address both of these issues, we re-analyzed the N250 amplitude using the mean amplitude in a ±30 ms time-window centered on the grand average N250 peak, calculated separately per

hemisphere, per session (≈306 ms). Additionally, we analyzed the P200 using the similar approach: a ±20 ms time-window centered on the grand average P200 peak (≈234 ms), separately per hemisphere, per session.

The new measure of N250 mean amplitude was submitted to a 2 *(hemisphere)* × 2 *(filter orientation)* × 5 *(bandwidth)* × 2 *(session)* ANOVA. Unlike what was found with our original measure of N250 amplitude, the main effects of *filter orientation* ($F_{(1,10)} = 3.67$, $p = 0.0845$, $\eta^2_G = 0.003$) and *bandwidth* ($F_{(4,40)} = 1.90$, $p_{HF} = 0.129$, $\eta^2_G = 0.004$) were not significant. Critically, the *filter orientation* × *bandwidth* interaction ($F_{(4,40)} = 3.50$, $p_{HF} = 0.0206$, $\eta^2_G = 0.005$) remained significant because the main effect of *bandwidth* was larger for vertically oriented filters ($F_{(4,40)} = 3.10$, $p_{HF} = 0.0259$, $\eta^2_G = 0.0116$) than for horizontally oriented filters ($F_{(4,40)} = 2.17$, $p_{HF} = 0.114$, $\eta^2_G = 0.010$). This interaction is similar to the one obtained with our original N250 measure. Furthermore, as the reviewer predicted, there was a significant main effect of *session* ($F_{(1,10)} = 11.3$, $p = 0.0072$, $\eta^2_G = 0.0139$), reflecting the fact that the N250 amplitude was greater during the first session ($-2.78\,\mu V$) than the second session (-1.94). The *filter orientation* × *bandwidth* × *hemisphere* interaction ($F_{(4,40)} = 2.64$, $p_{HF} = 0.0741$, $\eta^2_G = 0.0014$) was not significant. All other interactions did not approach significance ($F \leq 1.48$, $p \geq 0.25$, $\eta^2_G \leq 0.0017$).

P200 mean amplitude was submitted to a 2 *(hemisphere)* × 2 *(filter orientation)* × 5 *(bandwidth)* × 2 *(session)* ANOVA. There was a significant main effect of *filter orientation* ($F_{(1,10)} = 5.40$, $p = 0.0425$, $\eta^2_G = 0.0105$). The main effects of *bandwidth* ($F_{(4,40)} = 1.24$, $p_{HF} = 0.309$, $\eta^2_G = 0.003$), *hemisphere* ($F_{(1,10)} = 0.057$, $p = 0.816$, $\eta^2_G = 0.0005$), and *session* ($F_{(1,10)} = 2.93$, $p = 0.118$, $\eta^2_G = 0.0117$) were not significant. The *filter orientation* × *bandwidth* × *hemisphere* interaction was not significant ($F_{(4,40)} = 2.84$, $p_{HF} = 0.0559$, $\eta^2_G = 0.0015$). All other interactions also were not significant ($F \leq 1.49$, $p \geq 0.22$, $\eta^2_G \leq 0.0033$). We analyzed the main effect of *filter orientation* by comparing the average of the horizontally filtered ($M = 0.557\,\mu V$) and vertically filtered ($M = 1.23$) conditions to the full face condition ($M = 0.736$). Average P200 amplitude was larger (i.e., more positive) in the vertical filter condition than the horizontal filter condition ($t_{10} = 2.32$, $p = 0.0425$, $d = 0.701$), but the P200 to full faces did not differ from the P200 amplitude in the horizontal ($t_{10} = 0.609$, $p = 0.556$, $d = 0.184$) or vertical conditions ($t_{10} = 1.21$, $p = 0.255$, $d = 0.364$). Like the N250, the P200 was different for vertically filtered faces than either horizontally filtered or unfiltered faces. The P200 amplitude was not significantly modulated by bandwidth,

suggesting that the P200 was less sensitive than the N250 to facial structure.

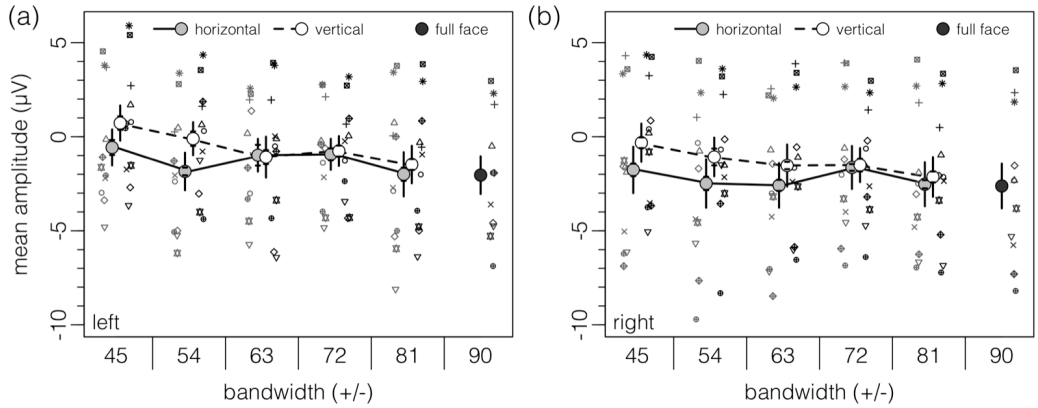

Figure 2.6 – N250 mean amplitude averaged across all subjects, separately for left and right hemispheres. Error bars represent ± 1 SEM, and within-subject corrected error bars are shown as horizontal ticks. Refer to figure 2.5 for details.

P100 amplitude & latency

The N170 is thought to be the earliest ERP component to be sensitive to faces *per se*. Differential activity before the N170 usually is thought to reflect differences in low-level stimulus characteristics (Johnson and Olshausen, 2003; VanRullen and Thorpe, 2001; Rousselet et al., 2007b). The orientation filtering that we used to construct our faces clearly produced low-level stimulus differences among our conditions (e.g., Table 2.1 & Figure 2.1a). These stimulus differences may have produced differences in very early ERP components (Figure 2.4), which in turn may have affected subsequent components (i.e., the N170). To investigate this possibility, we submitted P100 amplitudes to a 2 (*filter orientation*) × 5 (*bandwidth*) × 2 (*hemisphere*) ANOVA. The ANOVA failed to find any significant effects, and all of the effect sizes were very small (in all cases, $0.343 \leq F \leq 1.76$, $0.158 \leq p \leq 0.571$, $0.0001 \leq \eta^2_G \leq 0.0006$). This analysis indicates that the variability in P100 amplitude was not associated with changes in filter orientation or bandwidth, and that P100 amplitude was similar in the two hemispheres.

Despite not being significantly modulated by our independent variables, variability in the P100 amplitude could still be associated with the systematic variability we saw in the N170 amplitude. To test this idea, we re-analyzed the N170 amplitude results, but this time included P100 amplitude as a covariate. The ANCOVA revealed

significant main effects of *filter orientation* ($F_{(1,9)} = 18.3$, $p = 0.0021$, $\eta^2_G = 0.0069$) and *bandwidth* ($F_{(4,39)} = 8.07$, $p < 0.0001$, $\eta^2_G = 0.0213$). The main effect of *hemisphere* ($F_{(1,9)} = 0.649$, $p = 0.441$, $\eta^2_G = 0.0069$) was not significant. The *hemisphere* × *filter orientation* interaction ($F_{(1,9)} = 2.28$, $p = 0.166$, $\eta^2_G = 0.0004$) was not significant, but the *hemisphere* × *bandwidth* interaction ($F_{(4,39)} = 5.03$, $p = 0.0023$, $\eta^2_G = 0.0009$) was significant. The three-way *hemisphere* × *filter orientation* × *bandwidth* interaction ($F_{(4,39)} = 0.128$, $p = 0.971$, $\eta^2_G < 0.0001$) was not significant. Critically, the *filter orientation* × *bandwidth* interaction ($F_{(4,39)} = 5.87$, $p = 0.0008$, $\eta^2_G = 0.0055$) was significant. This interaction was analyzed by conducting separate ANCOVAs on the horizontal and vertical filter conditions. We found that the main effect of *bandwidth* was significant for the vertical filter orientation ($F_{(4,39)} = 11.2$, $p < 0.0001$, $\eta^2_G = 0.0453$), but not the horizontal filter orientation ($F_{(4,39)} = 2.17$, $p = 0.0902$, $\eta^2_G = 0.0065$). These analyses suggest that the N170 amplitude effects reported earlier are largely unaffected by incorporating P100 amplitude into the analyses.

We also re-analyzed the results obtained with our original and revised measure of N250 amplitude using the P100 amplitude as a covariate. For brevity, we report the ANCOVA results using the original measure of the N250, but the pattern of results was unchanged for the revised measure of the N250 amplitude. The main effects of *filter orientation* ($F_{(1,9)} = 12.5$, $p = 0.0064$, $\eta^2_G = 0.0118$) and *bandwidth* ($F_{(4,39)} = 4.69$, $p = 0.0035$, $\eta^2_G = 0.0208$) were significant, but the main effect of *hemisphere* ($F_{(1,9)} = 0.958$, $p = 0.353$, $\eta^2_G = 0.0102$) was not. The two-way interactions between *hemisphere* and *filter orientation* ($F_{(1,9)} = 0.169$, $p = 0.691$, $\eta^2_G < 0.0001$), and between *hemisphere* and *bandwidth* ($F_{(4,39)} = 1.94$, $p = 0.123$, $]eta^2_G = 0.0007$) were not significant, nor was the three-way interaction between *hemisphere*, *filter orientation*, and *bandwidth* ($F_{(4,39)} = 0.923$, $p = 0.460$, $\eta^2_G = 0.0007$). The *filter orientation* × *bandwidth* interaction ($F_{(4,39)} = 2.33$, $p = 0.0727$, $\eta^2_G = 0.0034$) was not significant. Nevertheless, as was done in our other analyses, we conducted separate ANCOVAs to evaluate the effect of *bandwidth* in the horizontal and vertical filter conditions. As was found in the analysis without a covariate, the main effect of *bandwidth* was significant for both the vertical filter orientation ($F_{(4,39)} = 5.27$, $p < 0.0017$, $\eta^2_G = 0.0366$) and the horizontal filter orientation ($F_{(4,39)} = 3.13$, $p = 0.0250$, $\eta^2_G = 0.0145$). These analyses suggest that incorporating P100 amplitude into our analyses of N250 amplitude slightly reduced the magnitude of the *filter orientation* × *bandwidth* interaction, but otherwise left the N250 amplitude effects reported earlier largely unaffected.

The P100 latency was also submitted to a $2 \times 5 \times 2$ ANOVA. The ANOVA found significant main effects of filter orientation ($F_{(1,10)} = 7.63$, $p = 0.0201$, $\eta^2_G = 0.0480$) and bandwidth ($F_{(4,40)} = 35.9$, $p_{HF} < 0.0001$, $\eta^2_G = 0.175$), but not *hemisphere* ($F_{1,10)} = 1.30$, $p = 0.281$, $\eta^2_G = 0.0012$). The two-way interactions between *filter orientation* × *bandwidth* ($F_{(4,40)} = 1.67$, $p = 0.176$, $\eta^2_G = 0.009$), *filter orientation* × *hemisphere* ($F_{(1,10)} = 2.72$, $p = 0.130$, $\eta^2_G = 0.0007$), and *bandwidth* × *hemisphere* ($F_{(4,40)} = 1.18$, $p = 0.333$, $\eta^2_G = 0.005$) were not significant, nor was the three-way interaction ($F_{(4,40)} = 0.0362$, $p = 0.997$, $\eta^2_G = 0.00006$).

Finally, P100 latency was used as a covariate in a re-analysis of the N170 latency in a $2 \times 5 \times 2$ ANCOVA, which found that the main effect of *bandwidth* ($F_{(4,39)} = 4.91$, $p = 0.0027$, $\eta^2_G = 0.0453$) remained significant, and the main effects of *filter orientation* ($F_{(1,9)} = 0.012$, $p = 0.916$, $\eta^2_G < 0.0001$) and *hemisphere* ($F_{(1,9)} = 0.064$, $p = 0.807$, $\eta^2_G = 0.0003$) remained non-significant. The two-way interaction between *filter orientation* × *bandwidth* ($F_{(4,39)} = 8.12$, $p < 0.0001$, $\eta^2_G = 0.0470$) was significant, and was due to larger effect of *bandwidth* in the vertical filter condition ($F_{(4,39)} = 11.7$, $p < 0.0001$, $\eta^2_G = 0.1256$) than the horizontal filter condition ($F_{(4,39)} = 1.79$, $p = 0.150$, $\eta^2_G = 0.0334$). The *filter orientation* × *hemisphere* interaction ($F_{(1,9)} = 12.4$, $p = 0.0065$, $\eta^2_G = 0.0040$) also was significant, and was driven by a larger main effect of *bandwidth* for the right hemisphere ($F_{(4,39)} = 11.1$, $p < 0.0001$, $\eta^2_G = 0.1215$) than the left hemisphere ($F_{(4,39)} = 2.57$, $p = 0.0528$, $\eta^2_G = 0.0318$). The *bandwidth* × *hemisphere* interaction ($F_{(4,39)} = 2.41$, $p = 0.0653$, $\eta^2_G = 0.0079$) and the three-way interaction ($F_{(4,39)} = 1.62$, $p = 0.190$, $\eta^2_G = 0.0020$) were not significant. The analysis suggests that incorporating P100 latency into our analyses of N170 latency revealed a *bandwidth* × *hemisphere* interaction but otherwise left the other effects, including the *filter orientation* × *bandwidth* interaction, unaffected.

Correlations between behaviour & ERPs

Previous studies have shown that relative sensitivity for horizontal compared to vertical facial structure is correlated with better performance in a face identification task and with the magnitude of the face inversion effect (Pachai et al., 2013). Here we examine whether a horizontal bias in our behavioural and ERP measures were correlated with face identification accuracy. We computed behavioural horizontal bias as the difference between accuracy in the H45 and V45 conditions, which are the two conditions that used the largest non-overlapping horizontal and vertical filters. Using the same formula,

ERP horizontal bias was also calculated for the P100 amplitude, P100 latency, N170 amplitude, N170 latency, and N250 amplitude, separately for each hemisphere. We correlated each of our measures of horizontal bias with identification accuracy to full, unfiltered faces.

The correlation between behavioural horizontal bias and full face identification (Figure 2.7a) was not significant ($r = 0.46$, $t_9 = 1.57$, $p = 0.076$, 1-tailed). Although the correlation was not statistically significant, its magnitude was similar to the value of 0.52 reported by Pachai et al. (2013) who had a significantly larger sample size ($n = 32$) than the one used in the current experiment, and who measured face identification *thresholds*, which had greater variability than our measure of proportion correct. Hence, although the correlation was not statistically significant, our results are consistent with previous reports that relative sensitivity to horizontal structure is correlated positively with, and accounts for 20-25% of the variance in, face identification performance.

The relations between response accuracy in the full face condition and horizontal bias of the P100, N170, and N250 are shown in Figure 2.7 b-f. As expected, full-face identification accuracy was not significantly correlated with either horizontal bias in P100 amplitude (Figure 2.7b) or P100 latency (Figure 2.7c). Full-face identification accuracy also was not correlated with the horizontal bias of N170 amplitude (Figure 2.7d) or N170 latency (Figure 2.7e). However, full-face identification accuracy *was* significantly correlated with the horizontal bias of N250 amplitude (Figure 2.7f) in both hemispheres (left: $r = 0.787$, $t_9 = -3.82$, $p = 0.0041$; right: $r = 0.666$, $t_9 = -2.68$, $p = 0.0253$). Averaging the results across hemispheres did not change the results: full face identification accuracy was not correlated with the horizontal bias of the P100 amplitude ($r = -0.308$, $t_9 = -0.97$, $p = 0.358$), P100 latency ($r = -0.407$, $t_9 = -1.34$, $p = 0.214$), N170 amplitude ($r = -0.169$, $t_9 = -0.513$, $p = 0.620$), or N170 latency ($r = 0.270$, $t_9 = 0.840$, $p = 0.423$), but was significantly correlated with the horizontal bias of the N250 amplitude ($r = -0.799$, $t_9 = -3.99$, $p = 0.003$).

Given that we re-analyzed the N250 using a reviewer-suggested time-window, and added a P200 analysis, we asked if either of these measures were correlated with full face identification. P200 and N250 were averaged across hemispheres, horizontal bias of the mean amplitudes was calculated, and the horizontal biases were correlated with full face identification accuracy separately in each session. Full face identification was not correlated with P200 horizontal bias in the first ($r = -0.349$, $t_9 = -1.12$, $p = 0.293$) or second ($r = 0.020$, $t_9 = 0.061$, $p = 0.953$) session, nor was it correlated

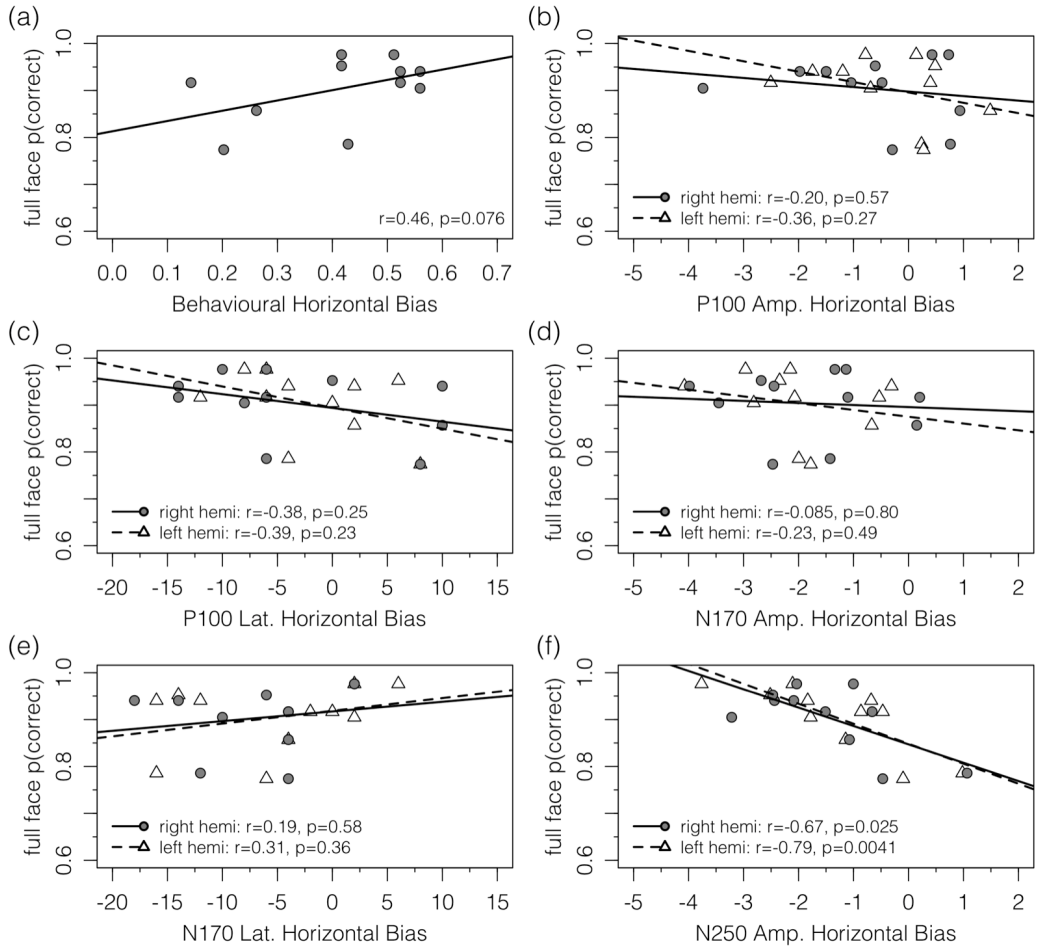

Figure 2.7 – Correlations between full face identification accuracy (y-axes) and different measures of horizontal bias (x-axes). Horizontal bias was measured for the **A)** behaviour, **B)** P100 amplitude, **C)** P100 latency, **D)** N170 amplitude, **E)** N170 latency, and **F)** N250 amplitude. Data are presented separately for the left (grey triangle, dashed line) and right (white circle, solid line) hemispheres. Pearson's correlations were computed using least squares method, with correlation values reported in the text. The only significant correlations were for the N250 mean amplitude (E) in both hemispheres.

with the revised N250 horizontal in the first ($r = -0.449$, $t_9 = -1.51$, $p = 0.166$) or second ($r = -0.312$, $t_9 = -0.981$, $p = 0.352$) session. However, the revised N250 horizontal bias averaged across sessions was significantly correlated with full face identification accuracy ($r = -0.685$, $t_9 = -2.82$, $p = 0.020$). Given this effect of session, we re-calculated the correlation between our original N250 measure (averaged across hemispheres) and behaviour for each session: as was found with the revised N250 measure, the correlation was smaller and non-significant in both the first ($r = -0.449$, $t_9 = -1.51$, $p = 0.166$) and second ($r = -0.311$, $t_9 = -0.981$, $p = 0.352$) session.

The fact that the brain-behaviour correlation depended significantly on the precise definition of the N250 raises the possibility that the significant correlation shown in Figure 2.7 was spurious and/or driven by a few, sparse time points in the ERP. To evaluate this idea, we correlated face identification accuracy and ERP-horizontal bias at every time-point from -200 to 500 ms, separately for the left and right hemispheres. We found that the correlation was significant for sustained periods of 250-285 ms and 264-288 ms in the right and left hemispheres, respectively (Figure 2.8). We also found, unexpectedly, a sustained correlation from 452-500 ms, but only in the right hemisphere.

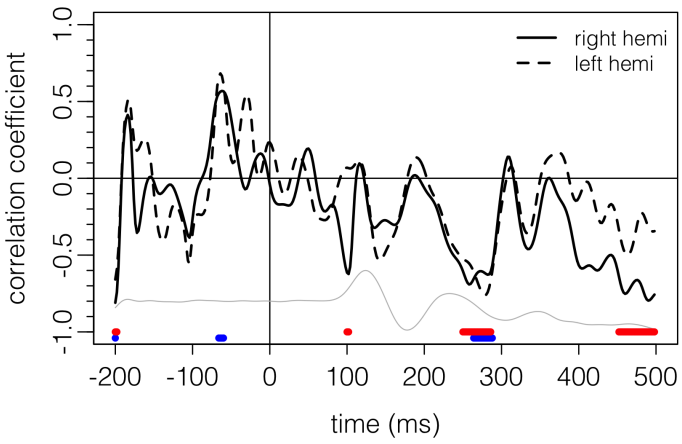

Figure 2.8 – Time-series of the Pearson's least-squares correlation coefficient (r) between neural horizontal bias (H45 - V45) and full face identification accuracy, separately for the right (solid) and left (dashed) hemispheres. Along the bottom, red and blue dots indicate statistically significant correlations ($p < 0.05$, uncorrected) for the right and left hemispheres, respectively. In both hemispheres, there was a sustained correlation between approximately 250-300 ms, which falls between the nearest peak and trough in the ERP. In the right hemisphere, there was a second sustained correlation between 450-500 ms. As a temporal reference, the grand-average ERP is shown in the thin grey line – note that the y-axis does not apply to this ERP. See Figure 2.4 for ERP traces.

2.5 Discussion

We measured response accuracy and ERPs in subjects performing a 1-of-6 identification task with faces that varied systematically in diagnostic vertical and horizontal structure. We found that identification was most accurate when faces contained horizontal structure, and was least accurate when faces lacked horizontal structure (Figure 2.2). Although the correlation between full face identification accuracy and our behavioural measure of horizontal tuning was not statistically significant (Figure 2.7a), the direction and magnitude of the correlation was similar to the one reported by Pachai et al. (2013). In addition, we found that N170 and N250 amplitudes were greatest (Figures 2.5 a&b and 2.6, respectively), and N170 latency was shortest (Figure 2.5 c&d), in response to faces that contained horizontal structure. Finally, we found that the horizontal bias of the N250 (Figure 2.7f), but not the N170 (Figure 2.7 d&e), was significantly correlated with response accuracy with unfiltered faces. Overall, our results are consistent with previous studies showing that horizontal facial structure is important for upright face identification (Dakin and Watt, 2009; Pachai et al., 2013, 2017), and suggest that the behavioural effects of manipulating oriented facial structure are more closely associated with the N250 than the N170.

We found that faces containing horizontal structure evoked the largest and earliest N170. Interestingly, the presence of horizontal structure had different effects on N170 amplitude and latency. Specifically, horizontal structure was necessary and sufficient to produce an N170 with an amplitude equal to that evoked by an unfiltered faces; however, horizontal structure was necessary but *not* sufficient to evoke an N170 with the same latency as the one evoked by an unfiltered face. Indeed, we found that, relative to the unfiltered condition, N170 latencies were longer in both the H45 and V45 conditions. N170 latency also was longer in the H81 condition, which contained structure at all orientations except those within ± 9 deg of vertical. These results are similar to those reported by Jacques et al. (2014), who found that phase randomization of vertical facial structure produced a slightly delayed N170, despite the presence of undistorted horizontal structure. In their study, all of the faces contained structure at all orientations, though the phase randomization reduced the information conveyed by selected bands of orientations. One interpretation of our results, and those of Jacques et al. (2014), is that N170 latency is more sensitive than N170 amplitude to image manipulations that render a face less *face-like*. According to this hypothesis,

manipulations of orientated structure that do not introduce obvious image distortions should have minimal affects on N170 latency. Hashemi et al. (2012) provide data that supports this idea: they used a 1-of-10 identification task with faces whose diagnostic orientation structure was manipulated using filters that were similar to those used here. However, instead of removing the filtered orientations, the filtered components were replaced with non-informative, oriented structure that was derived from the average of the 10 faces comprising the stimulus set. Hence, all faces in all conditions contained structure at all orientations and therefore appeared equally face-like in all conditions (for stimulus examples, see Pachai et al., 2018). When faces contained diagnostic horizontal structure but non-diagnostic vertical structure, Hashemi et al. (2012) found the N170 was not delayed relative to the N170 evoked by an unfiltered face. These results are consistent with the idea that N170 latency is more sensitive than N170 amplitude to manipulations, such as inversion (Bentin et al., 1996; Rossion et al., 2000; Itier et al., 2007; Rousselet et al., 2008b), that visibly distort a face: that is, the shortest latency N170 will be to an unmanipulated, intact, familiar face, and manipulations that produce visible differences from this baseline face will result in longer N170 latencies.

Unlike what was found with N170 latency, N170 amplitude in the H45 condition, which contained horizontal but not vertical structure, did not differ from N170 amplitude in the full-face condition. On the other hand, N170 amplitude in the V45 condition, which contained vertical but not horizontal structure, was significantly lower than N170 amplitude in the H45 and full-face conditions. Taken together, these results suggest that horizontal facial structure was necessary and sufficient to obtain normal N170 amplitudes. These results are consistent with Jemel et al.'s (2003) finding that N170 amplitude is approximately linearly related to the signal-to-noise ratio of full-face stimuli embedded in white noise if we assume that the effects of the noise were caused primarily by masking of horizontal facial structure (also see Rousselet et al., 2008b). Previous studies have shown that sensitivity to horizontal structure also is important for face discrimination and identification (Pachai et al., 2013; Goffaux and Greenwood, 2016), and therefore one might expect that N170 amplitude would be associated with behavioural measures of discrimination and identification accuracy. However, contrary to this hypothesis, we found a very small and non-significant correlation between the N170 amplitude and behaviour (Figure 2.7d). One potential explanation for this failure to find a correlation is that N170 amplitude depends on the presence of horizontal

facial structure regardless of whether that information is or is not informative about face identity. This hypothesis is consistent with results reported by Hashemi et al. (2012), who found that N170 amplitude in a face identification task was similar in conditions that contained informative and non-informative horizontal facial structure. Also, Jacques et al. (2014) found that scrambling the relative phase spectrum of horizontal facial structure had negligible effects on N170 amplitude evoked by upright faces. Both findings are consistent with the idea that N170 amplitude is not particularly sensitive to the information about face identity that may be conveyed by horizontal facial structure. Instead, the N170 may reflect neural processes that use horizontal structure to detect the presence of eyes (Nemrodov et al., 2014; Sekuler et al., 2015; de Lissa et al., 2014; Bentin et al., 1996; Itier et al., 2007; Rousselet et al., 2014).

Like the N170, the N250 amplitude was most negative for conditions containing horizontal structure (Figure 2.6). However, we also found a significant correlation between the horizontal bias of N250 amplitude and full-face identification accuracy (Figure 2.7f). This result suggests that N250 amplitude, unlike N170 amplitude, may be associated with *diagnostic* horizontal facial structure. This hypothesis is consistent with previous studies showing that the N250 is sensitive to the information that distinguishes a particular face from the average of an ensemble of faces (Zheng et al., 2012), and that the N250 responds differentially to familiar and unfamiliar faces (Schweinberger et al., 2002, 2004; Kaufmann et al., 2009; Tanaka et al., 2006). When the N250 was measured using a later time-window, the correlation was smaller but still significant. This result, together with the correlations at each time point in the ERP (Figure 2.8), suggests that the horizontal bias found between 250 and 300 ms after stimulus onset may not necessarily be constrained to correspond to the N250 *per se*, or to any other ERP component defined by a peak. Therefore, the so-called N250 may have varied significantly across subjects and therefore produced a grand average N250 that was unlike N250s in individual subjects. Even within subjects, the N250 varied between conditions and between sessions, such that the correlation with behaviour was smaller and not significant when the N250 was calculated separately in each session. Furthermore, initial studies of the N250 reported difficulty in finding a well-defined N250 for all subjects and conditions (Tanaka et al., 2006), and therefore subsequent studies have used several time-windows to calculate an average N250 amplitude. For instance, Tanaka et al. (2006) used 230-320 ms, Kaufmann et al. (2009) used 240-280 ms, and a recent review reports the N250 occurring between 200 ms to

400 ms (Schweinberger and Neumann, 2016). In fact, as part of a cognitive model of face perception, activity starting at 450 ms onwards is thought to reflect high level semantic representations and name retrieval (Schweinberger and Neumann, 2016), both of which would intrinsically rely on the identity processing which occurs earlier during the N250 time. In light of the variability in and correlations with the later parts of the ERP, analyses that are less reliant on pre-defined times-of-interest, such as the one illustrated in Figure 2.8, or robust approaches such as that seen in Rousselet et al. (2014), will be useful to investigate the association between neural horizontal bias in the entire ERP and face identification.

In summary, the current study shows that manipulating the horizontal and vertical structure of upright faces in a 1-of-6 face identification task similarly affects response accuracy, N170 amplitude and latency, and N250 amplitude. We found that the presence of horizontal facial structure was associated with higher response accuracy, higher amplitude and shorter latency N170s, and higher amplitude N250s. Finally, we found that the identification accuracy for unfiltered faces was correlated with the horizontal bias of N250 amplitude but not N170 amplitude. We suggest that both the N170 and N250 are sensitive to the presence of horizontal facial structure, but only the N250 is affected by the information value of such structure for face identification.

References

Amihai, I., Deouell, L. Y., and Bentin, S. (2011). Neural adaptation is related to face repetition irrespective of identity: A reappraisal of the N170 effect. *Experimental Brain Research*, 209(2):193–204.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–84.

Balas, B. J., Huynh, C., Saville, A., and Schmidt, J. (2015a). Orientation biases for facial emotion recognition during childhood and adulthood. *Journal of Experimental Child Psychology*, 140:171–83.

Balas, B. J., Schmidt, J., and Saville, A. (2015b). A face detection bias for horizontal orientations develops in middle childhood. *Frontiers in Psychology*, 6(June):1–8.

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, 8(6):551–65.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–6.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, pages 2–5.

Dakin, S. and Watt, R. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

de Haan, M., Pascalis, O., and Johnson, M. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, 14(2):199–209.

de Heering, A., Goffaux, V., Dollion, N., Godard, O., Durand, K., and Baudouin, J. Y. (2016). Three-month-old infants' sensitivity to horizontal information within faces. *Developmental Psychobiology*, 58(4):536–42.

de Lissa, P., McArthur, G., Hawelka, S., Palermo, R., Mahajan, Y., and Hutzler, F. (2014). Fixation location on upright and inverted faces modulates the N170. *Neuropsychologia*, 57(1):1–11.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21.

Diamond, R. and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2):107–17.

Eimer, M. (2000). Effects of face inversion on the structural encoding and recognition of faces: Evidence from event-related brain potentials. *Cognitive Brain Research*, 10:145–58.

Farah, M., Wilson, K., Drain, M., and Tanaka, J. W. (1998). What is "special" about face perception? *Psychological Review*.

Gaspar, C. M., Sekuler, A. B., and Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research*, 48(28):2817–26.

Gauthier, I. and Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, 103:322–30.

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

Goffaux, V. and Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, 1(September):143.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., and Goebel, R. (2016). Horizontal tuning for faces originates in high-level fusiform face area. *Neuropsychologia*, pages 1–11.

Goffaux, V. and Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, 6:34204.

Goffaux, V., Poncin, A., and Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 Year Old). *PLoS ONE*, 10(9):1–17.

Gold, J., Bennett, P. J., and Sekuler, A. B. (1999a). Signal but not noise changes with perceptual learning. *Nature*, 402(6758):176–8.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999b). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39:3537–60.

Gold, J. M., Sekuler, A. B., and Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28(2):167–207.

Hashemi, A., Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2012). Exploring the relationship between the N170 inversion effect and horizontal tuning. *Journal of Vision*, 12(9):624.

Heisz, J. J., Watter, S., and Shedden, J. M. (2006). Automatic face identity encoding at the N170. *Vision Research*, 46(28):4604–14.

Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2007). Inverting houses and textures: Investigating the characteristics of learned inversion effects. *Vision Research*, 47:3350–59.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, 49(18):2273–84.

Huynh, C. M. and Balas, B. J. (2014). Emotion recognition (sometimes) depends on horizontal orientations. *Attention, Perception & Psychophysics*.

Itier, R. J., Alain, C., Sedore, K., and McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience*, 19(11):1815–26.

Itier, R. J. and Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *NeuroImage*, 15(2):353–72.

Jacques, C. and Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *NeuroImage*, 36:863–76.

Jacques, C., Schiltz, C., and Goffaux, V. (2014). Face perception is tuned to horizontal orientation in the N170 time window. *Journal of Vision*, 14:1–18.

Jemel, B., Schuller, A.-M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., and Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *Neuroreport*, 14(16):2035–39.

Johnson, J. S. and Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7):499–512.

Kaufmann, J. J. M., Schweinberger, S. R., Burton, A. M., and Kaufmann, M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience*, 21(4):625–41.

Keil, M. S., Lapedriza, A., Masip, D., and Vitria, J. (2008). Preferred spatial frequencies for human face processing are associated with optimal class discrimination in the machine. *PLoS ONE*, 3(7):1–5.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., and Others (2007). What's new in Psychtoolbox-3. *Perception*, 36(14):1.

Konar, Y., Bennett, P. J., and Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, 21(1):38–43.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(NOV):1–12.

Loftus, G. R. and Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4):476–90.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press, Cambridge, Mass.

Mathworks (2007). MATLAB version 7.5 (R2007b).

Maurer, D., Grand, R., and Mondloch, C. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6):255–60.

Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2):61–64.

Murphy, J. and Cook, R. (2017). Revealing the mechanisms of human face perception using dynamic apertures. *Cognition*, 169:25–35.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39(23):3824–33.

Nemrodov, D., Anderson, T., Preston, F., and Itier, R. (2014). Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *NeuroImage*, 97:81–94.

Obermeyer, S., Kolling, T., Schaich, A., and Knopf, M. (2012). Differences between Old and Young Adults' Ability to Recognize Human Faces Underlie Processing of Horizontal Information. *Frontiers in Aging Neuroscience*, 4(April):3.

Olejnik, S. and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, 8(4):434–47.

Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2018). The bandwidth of diagnostic horizontal structure for face identification. *Perception*, 0(0):030100661875447.

Pachai, M. V., Bennett, P. J., Sekuler, A. B., Corrow, S., and Barton, J. J. S. (2015). Sensitivity to Horizontal Structure and Face Identification in Developmental Prosopagnosia and Healthy Aging. *Canadian Journal of Experimental Psychology*, 69(4):341.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to Information Conveyed by Horizontal Contours is Correlated with Face Identification Accuracy. *Frontiers in Psychology*, 4(74).

Pachai, M. V., Sekuler, A. B., Bennett, P. J., Schyns, P. G., and Ramon, M. (2017). Personal familiarity enhances sensitivity to horizontal structure during processing of face identity. *Journal of Vision*, 17(6):5.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics Transforming numbers into movies. *Spatial Vision*, 10(4):437–42.

Pernet, C. R., Chauveau, N., Gaspar, C. M., and Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational intelligence and neuroscience*, 2011:831409.

Piepers, D. W. and Robbins, R. A. (2012). A review and clarification of the terms "holistic,""configural,"and "relational"in the face perception literature. *Frontiers in Psychology*, 3:559.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rodger, H., Kelly, D., Blais, C., and Caldara, R. (2009). Inverting faces does not abolish cultural diversity in eye movements. *Perception*, 39(11):1491–1504.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., and Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11(1):69–74.

Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2007a). Single-Trial EEG Dynamics of Object and Face Visual Processing. *NeuroImage*, 36:843–62.

Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008a). Time course and robustness of ERP object and face differences. *Journal of Vision*, 8:1–18.

Rousselet, G. A., Ince, R. A. A., Rijsbergen, N. J. V., and Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *Journal of Vision*, 14(13):1–24.

Rousselet, G. A., Macé, Marc J.-M., Thorpe, S. J., and Fabre-Thorpe, M. (2007b). Limits of event-related potential differences in tracking object processing speed. *Journal of Cognitive Neuroscience*, 19(8):1241–58.

Rousselet, G. A., Macé, M., and Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision*, 4:13–21.

Rousselet, G. A., Pernet, C. R., Bennett, P. J., and Sekuler, A. B. (2008b). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, 9:98.

Royer, J., Willenbockel, V., Blais, C., Gosselin, F., Lafortune, S., Leclerc, J., and Fiset, D. (2017). The influence of natural contour and face size on the spatial

frequency tuning for identifying upright and inverted faces. *Psychological Research*, 81(1):13–23.

Schweinberger, S. R., Huddy, V., and Burton, A. M. (2004). N250r: a face-selective brain response to stimulus repetitions. *NeuroReport*, 15(9):1501–05.

Schweinberger, S. R. and Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex*, 80:141–153.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., and Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, 14(3):398–409.

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., and Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, 1210:204–15.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–6.

Sekuler, A. B., Pachai, M. V., Hashemi, A., and Bennett, P. J. (2015). Effects of size, fixation location, and inversion on face identification. *Journal of Vision*, 15(12):694.

Tanaka, J. W., Curran, T., Porterfield, A. L., and Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, 18(9):1488–97.

Tanaka, J. W. and Farah, M. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46A(2):225–45.

Tanskanen, T., Näsänen, R., Montez, T., Päällysaho, J., and Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cerebral Cortex*, 15(5):526–34.

Taubert, J., Goffaux, V., Van Belle, G., Vanduffel, W., and Vogels, R. (2016). The impact of orientation filtering on face-selective neurons in monkey inferior temporal cortex. *Scientific reports*, 6:21189.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87:154–63.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, 79:471–91.

VanRullen, R. and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4):454–61.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., Bub, D. N., and Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of experimental psychology. Human perception and performance*, 36(1):122–35.

Williams, C. and Henderson, J. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition*, 35(8):1977–85.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1):141–45.

Yu, D. and Chung, S. S. T. L. (2011). Critical orientation for face identification in central vision loss. *Optometry and vision science : official publication of the American Academy of Optometry*, 88(6):724–32.

Zheng, X., Mondloch, C. J., and Segalowitz, S. J. (2012). The timing of individual face recognition in the brain. *Neuropsychologia*, 50(7):1451–61.

Chapter 3

The behavioural and neural horizontal bias of older adults

3.1 Abstract

Horizontal structure is important for face identification, and recent studies have shown that younger adults have heightened behavioural and neural sensitivity to horizontal relative to vertical facial structure (i.e., horizontal bias). Also, older adults make more errors in face identification, and show a decreased horizontal bias. However, it is unclear if face identification in older adults is still associated with horizontal bias, or if they rely on an alternative strategy to identify faces. Specifically, do age-related changes in face identification translate to similar changes in the neural sensitivity to horizontal bias? Here, we recorded ERPs and behavioural accuracy from older adult observers (M = 73 years) performing the same task as Chapter 2: a 1-of-6 face identification task in conditions that parametrically manipulated the orientation structure of upright faces. Faces were filtered with ideal orientation filters centred on either 0 (horizontal) or 90 deg (vertical). Filter bandwidth was varied across conditions from ±45 to ±90 deg in steps of ±9 deg. Relative to the younger adults in Chapter 2, face identification accuracy was overall lower in older adults. Similar to younger adults, response accuracy was significantly higher for faces that contained horizontal structure than vertical structure, and behavioural horizontal bias was correlated with face identification accuracy. Like younger adults, N250 horizontal bias was correlated with face identification accuracy. Unlike younger adults, N170 amplitude horizontal bias also was correlated with face

identification accuracy. We suggest that face identification in older adults still relies on horizontal bias, but that relative to younger adults, additional factors such as inefficiencies in early visual processes have a greater effect on face identification.

3.2 Introduction

Identifying faces quickly and accurately is an important, everyday visual task, which most adults perform effortlessly. Most adults perceive upright faces, but not inverted faces, effortlessly, and this so-called face inversion effect presumably reflects the considerably greater experience – and, consequently, perceptual expertise – that most humans have discriminating and identifying upright faces (Valentine, 1988). One widely-held hypothesis is that experience with upright faces allows observers to develop processes that are sensitive to configural/holistic aspects of faces (Tanaka and Farah, 1993; Farah et al., 1998; Maurer et al., 2002). However, it has proven difficult to clearly define configural processing (Piepers and Robbins, 2012). Furthermore, several studies have shown that accuracy in face identification tasks are not associated with common measures of configural processing (Konar et al., 2010; Wang et al., 2012). Finally, the idea that configural processing contributes to the perception of upright faces, but not inverted faces, has been challenged by studies showing that inverted faces are processed similarly to upright faces, just less efficiently (Sekuler et al., 2004; Riesenhuber et al., 2004; Williams and Henderson, 2007; Gaspar et al., 2008; Willenbockel et al., 2010; Pachai et al., 2013; Murphy and Cook, 2017).

Recent attempts to characterize expertise have focussed on the importance of horizontal structure (see Figure 3.1 for an example of horizontal structure). Dakin and Watt (2009) demonstrated that several aspects of face perception are particularly sensitive to horizontal facial structure. Horizontal structure is necessary for the face inversion effect (Goffaux and Dakin, 2010). Importantly, individual differences in how much observers rely on horizontal structure relative to vertical structure (i.e., horizontal bias) when identifying faces is correlated with response accuracy and with the size of the inversion effect (Pachai et al., 2013; Chapter 2). That is, the better you are at using horizontal structure of upright faces, the larger your advantage in identifying upright relative to inverted faces. Furthermore, horizontal bias is associated with differences in face perception across ages and groups: For example, horizontal bias increases throughout childhood (Balas et al., 2015; Goffaux et al., 2015), decreases in older

adults (Obermeyer et al., 2012; Goffaux et al., 2015), and is smaller in individuals with prosopagnosia (Pachai et al., 2015). Finally, horizontal bias is larger for familiar than unfamiliar faces (Pachai et al., 2017), Hence, behavioural measures of horizontal bias have thus far provided a consistent narrative in which the efficient use of horizontal structure is an essential component, if not the critical mechanism, of expert face perception.

Neuroimaging studies have recently begun investigating the role of oriented structure in evoking face-related brain activity. For instance, the N170 event-related potential (ERP) is larger for faces than objects (Bentin et al., 1996; Rossion et al., 2000; Rousselet et al., 2004), but the face-related enhancement of the N170 is significantly attenuated when horizontal structure, but not vertical structure, is removed (Chapter 2). The N170 has a longer latency for inverted than upright faces (Bentin et al., 1996; Rossion et al., 2000), but only for faces with intact horizontal structure (Jacques et al., 2014). Furthermore, the human fusiform face area (FFA; Kanwisher et al., 1997), which exhibits activation that is correlated with the N170 (Deffke et al., 2007; Sadeh et al., 2010), is preferentially activated by horizontal compared to vertical facial structure (Goffaux et al., 2016). A slightly later ERP component, the N250 (Tanaka et al., 2006), also exhibits a bias for horizontal facial structure (Chapter 2). Importantly, variation across individuals in the horizontal bias of the N250, but not the N170, is correlated with face identification accuracy (Chapter 2), which is consistent with the idea that the N170 and N250 are associated with different aspects of face processing. Nevertheless, these results demonstrate that horizontal structure is important for face-related brain responses.

Several measures of face perception change during normal, healthy ageing. For example, ageing is associated with a reduction in the ability to identify faces (Grady et al., 1994; Konar et al., 2013), a reduction in horizontal bias for facial structure (Obermeyer et al., 2012), and a reduction in the face inversion effect (Chaby et al., 2011). Also, the N170 in older adults, relative to younger adults, has an enhanced amplitude (Wiese et al., 2008) and longer latency (Rousselet et al., 2009). Furthermore, in older adults the N170 generally is less selective for facial structure (Rousselet et al., 2010), and the N250 is less sensitive to face identity (Wiese et al., 2008). Hence, several behavioural and ERP measures of face discrimination and identification change during ageing.

It is not yet known if the face perception changes during normal, healthy ageing can

be recovered. If face perception expertise is achieved through mere experience seeing upright faces, then it is plausible that a perceptual learning task in which older adults identify faces may recover age-related deficits. We address this possibility in Chapter 4, and further ask if their improvements are due to changes in horizontal bias. However, before attempting to train older adults, we wanted to better understand their neural response to horizontal structure. So far, we know that older adults have: 1) difficulties in face perception (Grady et al., 1994; Konar et al., 2013), 2) a weaker horizontal bias (Obermeyer et al., 2012), 3) an enhanced (Wiese et al., 2008), delayed (Rousselet et al., 2009), and less face-selective (Rousselet et al., 2010) N170, and 4) an N250 that is less sensitive to face identity (Wiese et al., 2008). These effects appear to be theoretically linked, but they have not been experimentally tested in a single experiment. We aimed to fill this gap by testing older adults in a face identification task during which face stimuli were filtered to preserve different bandwidths of horizontally and vertically oriented structures, while simultaneously recording EEG. As in our work with younger adults (Chapter 2), we measured behavioural, N170, and N250 horizontal biases, and correlated each with full face identification accuracy. We were interested in whether age-related deficits in face perception can be linked to deficits in the neural markers of a horizontal bias.

3.3 Methods

3.3.1 Subjects

Nine older adults (3 male; range = 64-84 years old, M=73, SD=7.23) from the Hamilton area participated in the experiment. All were right handed, except for one female participant who was ambidextrous. All participants had normal or corrected-to-normal Snellen acuity, normal Pelli-Robson contrast sensitivity (range = 1.35-1.95, M=1.78, SD=0.20), and normal scores on the MoCA (range=26-29, M=27.7, SD=1.58; Nasreddine et al., 2005) and MMSE (range=29-30, M=29.8, SD=0.44; Folstein et al., 1983).

3.3.2 Apparatus & Stimuli

The experimental apparatus, methods of stimulus generation, and experimental conditions (Figure 3.1A) were identical to those used in Chapter 2, so only a brief summary

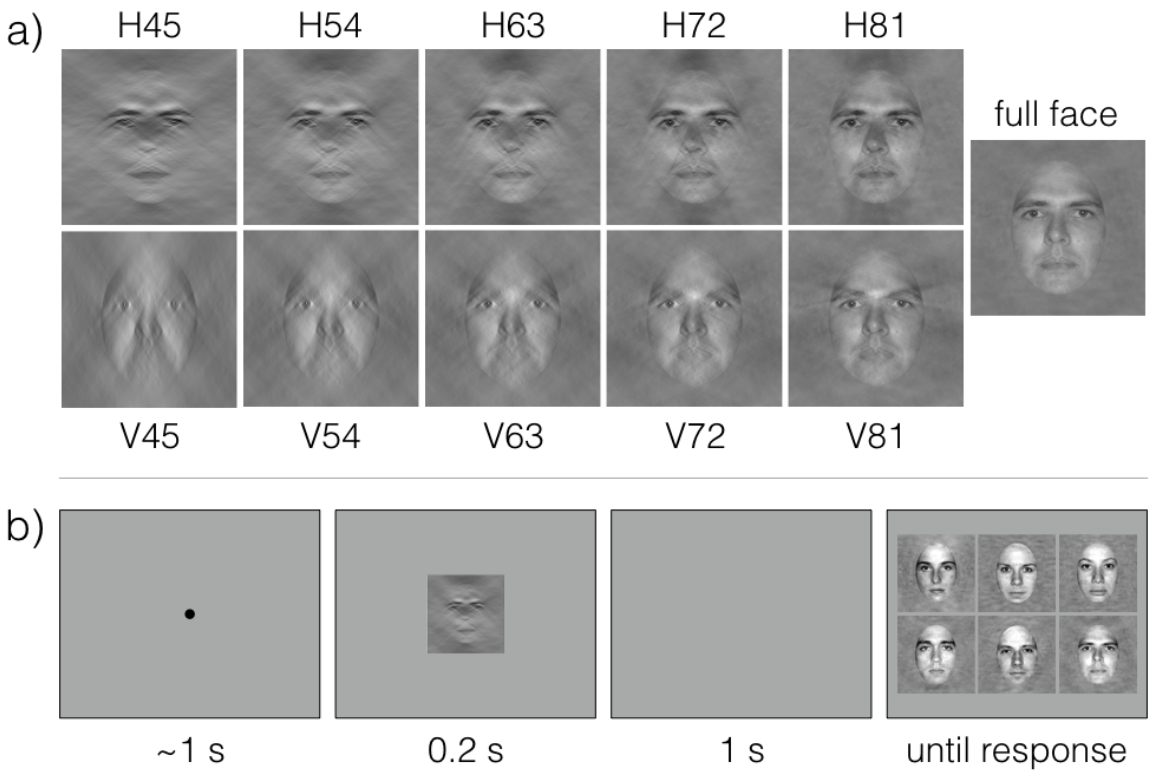

Figure 3.1 – **A)** Example stimuli of the 11 different conditions in Experiment 1. **B)** Visualization of the trial structure. Response screen alternatives were always presented in full face form.

is provided here. Six faces (3 male) from Gold et al. (1999) were each filtered with an ideal band-pass orientation filter centered on either 0 deg (horizontal) or 90 deg (vertical). Filter bandwidth varied from ±45 deg to ±81 deg in steps of ±9 deg, resulting in 10 filter conditions: five different filter bandwidths centered on horizontal, and five different filter bandwidths centered on vertical. In an eleventh condition, the filter bandwidth was ±90 deg, which covers all orientations: we refer to this as the 'full face' condition.

3.3.3 Procedure & Design

Stimuli were generated and responses were collected using MATLAB (MathWorks, 2007) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Each trial began with a central fixation point that was displayed for 900-1100 ms, followed by a 200 ms of a target face and, after a 1 s blank interval, a six-item response

screen (Figure 3.1B). The task was to identify the target face by clicking on one item in the response screen with a computer mouse. The response screen always consisted of the same six unfiltered faces presented in the same spatial arrangement. The target stimulus was selected randomly from the 11 filter conditions with the constraint that each condition occurred with equal frequency. Each combination of six face identities and 11 filter conditions was repeated 14 times, for a total of 924 trials spread across two sessions. Data were combined across all face identities but separated by condition, yielding 84 trials per condition.

3.3.4 Electrophysiology

EEG data were acquired during the behavioural task using a 256-channel HydroCel Geodesic Sensor Net (Electric Geodesics Inc., Eugene, Oregon; Tucker, 1993). Data were referenced online to electrode *Cz*, and sampled at 500 Hz. Data were then epoched from -200 ms to 998 ms relative to stimulus-onset, after which all post-processing was done in MATLAB using *EEGLAB* (Delorme and Makeig, 2004) using the same in-house pipeline as Chapter 2. ERP waveforms for the regions of interest for each subject and condition were then exported for plotting and statistical analysis in R (R Core Team and R Development Core Team, 2017). Both correct and incorrect response trials were used in forming the ERPs.

3.3.5 EEG Analysis

In general, ERP components such as the P100, N170, and N250 were not as clearly defined in older adults compared to the younger adults in Chapter 2. To estimate ERP amplitude and latency measures, we first computed the grand average ERP, separately in each hemisphere, across all conditions and subjects for two clusters of fifteen electrodes each centred on PO7 and PO8 (see Chapter 2, Figure 3 to see electrode cluster location), which are the electrodes that contain the largest face-related evoked potentials (Bentin et al., 1996; Rossion et al., 2000). We identified three peaks from the grand average ERP that were used as the centres of three ±20 ms time-windows which were then used to calculate P100, N170, and P200 amplitudes. The latencies of the peaks, which did not differ between hemispheres, were 144 ms, 206 ms, and 316 ms for the P100, N170, and P200, respectively. The N250 amplitude was calculated using a ±25 ms time-window centred on 410 ms, which is consistent with the time window

used in previous studies that measured the N250 in older adults (e.g., Wiese et al., 2008, used 285 ms and 410 ms for younger and older adults, respectively), and was confirmed by visual inspection of the ERP to occur slightly after the P200 peak. N170 latency was measured using the 50% area latency method (Luck, 2005) in the same time window used for measuring N170 amplitude. The 50% area latency is the latency at which the integral of the curve is split in half; here, the curve used to calculate the N170 latency was the absolute ERP waveform between 186 ms to 226 ms.

3.4 Results

3.4.1 Behavioural Results

Statistical analyses were performed with R (R Core Team and R Development Core Team, 2017). To maintain an orthogonal design, the full face condition was omitted from all analyses of variance (ANOVAs), unless otherwise noted. Instead, filtered conditions were compared to the full face condition using paired t tests, and a Bonferonni corrected α was used when appropriate. For all within-subjects effects with more than one degree-of-freedom, including the five level factor *bandwidth*, we did not assume sphericity and instead report Huynh-Feldt corrected p-values. Effect size was measured using Cohen's d for t tests, and generalized eta squared (η^2_G) for ANOVA factors, as described by Bakeman (2005) and Olejnik and Algina (2003) for repeated-measure designs. Lakens (2013) suggests that Cohen (1988)'s suggested benchmarks of effect size can be used when interpreting η^2_G (small: 0.01, medium: 0.04, large: 0.16).

Response accuracy is plotted as a function of filter condition in Figure 3.2A. Response accuracy was higher for the horizontal filter orientation relative to the vertical filter orientation, and accuracy increased as filter bandwidth increased for both filter orientations. The difference between horizontal and vertical filter orientations remained approximately similar across all bandwidths less than 90 deg. We analyzed the proportion of correct responses using a repeated-measures ANOVA that incorporated the factors *filter orientation* (horizontal and vertical) and *bandwidth* (± 45, 54, 63, 72, and 81 deg). The main effects of *filter orientation* ($F_{(1,8)} = 20.0$, $p = 0.002$, $\eta^2_G = 0.358$) and *bandwidth* ($F_{(4,32)} = 38.6$, $p_{HF} < 0.001$, $\eta^2_G = 0.331$) were significant. The *filter orientation* × *bandwidth* interaction ($F_{(4,32)} = 1.08$, $p_{HF} = 0.382$, $\eta^2_G = 0.009$) was not significant, indicating that the difference between the horizontal and vertical filter

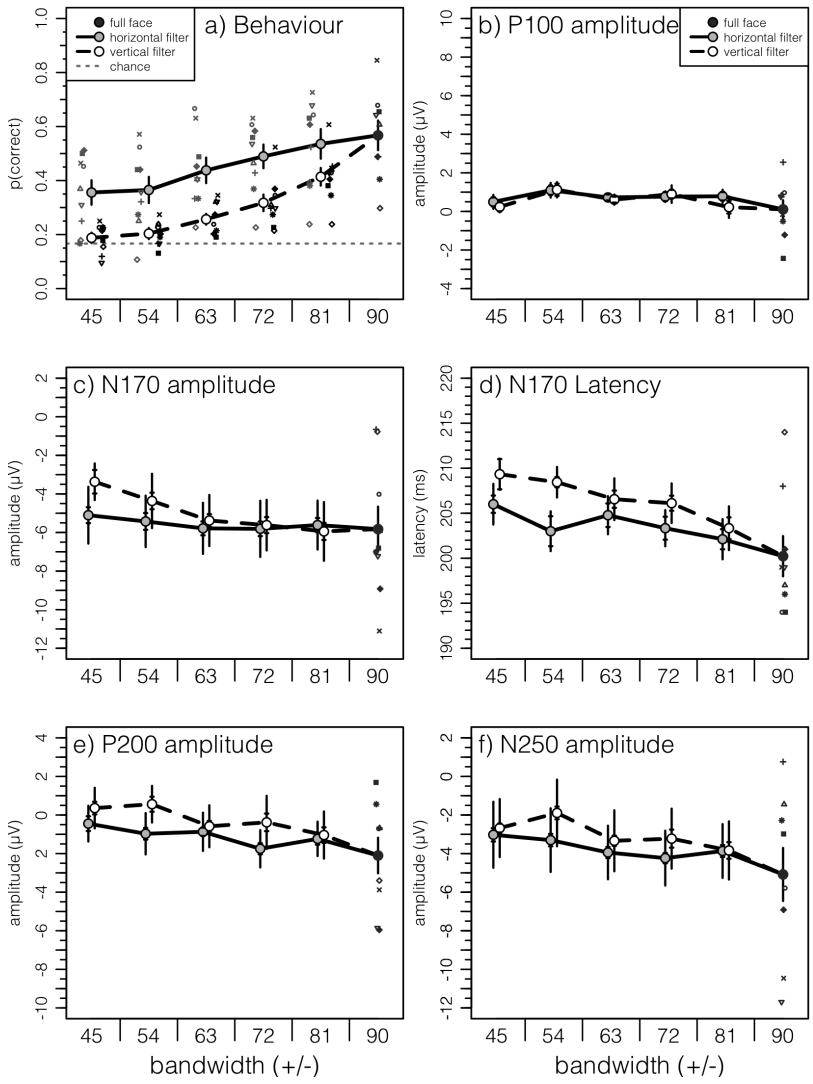

Figure 3.2 – A summary of behavioural and ERP results. **A)** Response accuracy plotted for each condition, with individual subjects depicted as smaller symbols. Dotted horizontal line represents chance performance. ERP measures are plotted in B-F and all are averaged across hemispheres: **B)** P100 mean amplitude, **C)** N170 mean amplitude, **D)** N170 50% area latency, **E)** P200 mean amplitude, and **F)** N250 mean amplitude. Error bars are +/-1 standard error of the mean (SEM), and horizontal ticks on error bars are within-subject corrected error bars, calculated specifically for repeated-measures designs (Cousineau, 2005; Morey, 2008). Individual subject data are presented for all conditions in the behaviour plot, and for aesthetics, for only the full face condition in the ERP plots to give a sense of the variability in each measure.

orientations remained approximately constant across bandwidths from 45 to 81 deg (Figure 3.2A).

We also compared accuracy in each filter condition to accuracy in the full face condition using two-tailed paired t tests. These contrasts are especially informative if we assume that response accuracy to full faces is an estimate of the upper limit of face identification accuracy in the current paradigm. We computed 10 comparisons, and therefore used a Bonferonni corrected per comparison α_{PC}=0.005 to maintain a family-wise α_{FW}=0.05. Response accuracy to full faces ($M_{FF} = 0.567$) was significantly higher than accuracy in the horizontal filter orientation conditions with the three smallest bandwidths ($M_{H45} = 0.356$, $t_8 = 5.34$, $p < 0.001$, $d = 1.78$; $M_{H54} = 0.365$, $t_8 = 6.54$, $p < 0.001$, $d = 2.18$; $M_{H63} = 0.438$, $t_8 = 4.66$, $p = 0.0016$, $d = 1.55$), but not the two largest bandwidths: ($M_{H72} = 0.489$, $t_8 = 2.79$, $p = 0.023$, $d = 0.93$; $M_{H81} = 0.536$, $t_8 = 0.99$, $p = 0.350$, $d = 0.33$). However, response accuracy to full faces was significantly higher than accuracy in the vertical filter orientation conditions at all bandwidths ($M_{V45} = 0.188$, $t_8 = 7.26$, $p < 0.001$, $d = 2.42$; $M_{V54} = 0.204$, $t_8 = 6.85$, $p < 0.001$, $d = 2.28$; $M_{V63} = 0.257$, $t_8 = 6.86$, $p < 0.001$, $d = 2.29$; $M_{V72} = 0.317$, $t_8 = 6.18$, $p < 0.001$, $d = 2.06$; $M_{V81} = 0.414$, $t_8 = 4.88$, $p = 0.001$, $d = 1.63$).

Lastly, we computed a measure of horizontal bias by subtracting V45 response accuracy from H45 response accuracy, and correlated it with full face identification accuracy (Figure 3.4A). Horizontal bias was significantly greater than zero ($M = 0.168$, $t_8 = 4.07$, one-tailed $p = 0.0018$, $d = 1.98$), and, as was found by Pachai et al. (2013) and in Chapter 2, it was significantly correlated with identification accuracy to full faces ($r = 0.637$, $t_7 = 2.19$, one-tailed $p = 0.0324$).

3.4.2 ERP Results

Visual inspection of the grand average ERPs (Figure 3.3A-C) suggests that orientation-related modulations of the ERP appeared larger for vertically oriented filters than horizontally oriented filters, indicating that the ERP to full faces is driven mostly by the presence of horizontal facial structure. Indeed, pairwise t tests performed at every time-point (Figure 3.3D-F) show that the ERP to full faces differed most from the ERP to V45-filtered stimuli (Figure 3.3F) than the ERP to H45-filtered stimuli (Figure 3.3E). Note that we depict ERP results only from the right hemisphere for brevity, but the pattern of results was mostly the same in the left hemisphere. All omnibus ANOVAs include hemisphere as a factor.

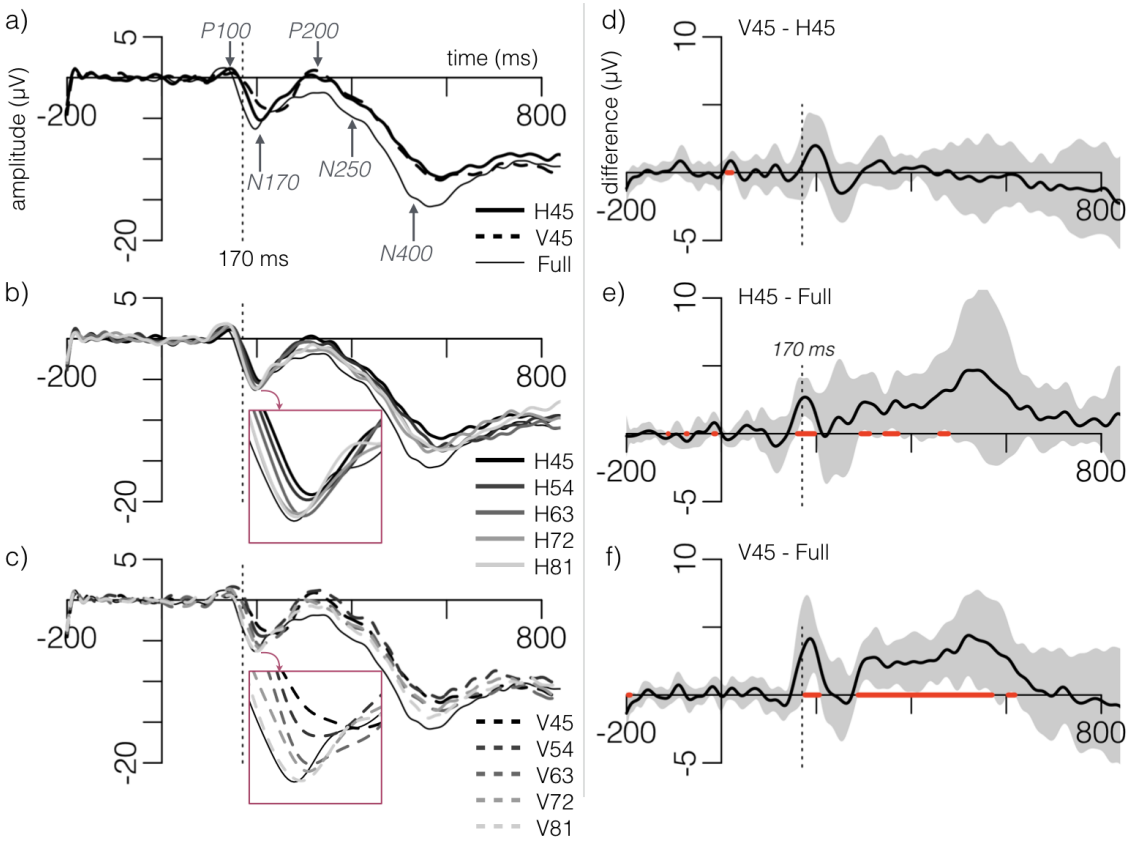

Figure 3.3 – Left: Grand average ERP waveforms for the right hemisphere for the full face condition (solid thin black line) paired with **A)** H45 and V45, **B)** all horizontal filter conditions, and **C)** all vertical filter conditions The N170 peak is specifically zoomed in B and C to better demonstrate the bandwidth related modulation. Right: Grand average ERP differences between **D)** V45 and H45, **E)** H45 and Full face, and **F)** V45 and Full face. Shaded region represents 95% confidence interval, and red along the x-axis represents significantly different than 0 at that time point ($p<0.05$, uncorrected)

3.4.2.1 P100 Amplitude

P100 mean amplitude was approximately constant across conditions (Figure 3.2B). P100 amplitudes were submitted to a 2 (*filter orientation*) × 5 (*bandwidth*) × 2 (*hemisphere*) repeated-measures ANOVA. The main effects of *filter orientation* ($F_{(1,8)} = 0.44$, $p = 0.52$, $\eta^2_G = 0.005$) and *bandwidth* ($F_{(4,32)} = 1.89$, $p_{HF} = 0.177$, $\eta^2_G = 0.051$), and the *orientation* × *bandwidth* interaction ($F_{(4,32)} = 0.44$, $p_{HF} = 0.78$, $\eta^2_G = 0.012$) were not significant. The main effect of *hemisphere* ($F_{(1,8)} = 4.82$, $p = 0.059$, $\eta^2_G = 0.26$) was qualified by a significant *hemisphere* × *filter orientation* interaction ($F_{(1,8)} = 11.1$, $p = 0.010$, $\eta^2_G = 0.004$), which was driven by a non-significant but larger main effect of *filter orientation* in the right ($F_{(1,8)} = 1.78$, $p = 0.219$, $\eta^2_G = 0.055$) than left ($F_{(1,8)} < 0.01$, $p = 0.95$, $\eta^2_G < 0.001$) hemisphere. The *hemisphere* × *bandwidth* interaction ($F_{(4,32)} = 0.34$, $p_{HF} = 0.85$, $\eta^2_G = 0.001$), and the three-way interaction ($F_{(4,32)} = 0.52$, $p_{HF} = 0.72$, $\eta^2_G = 0.004$) were not significant.

We computed P100 horizontal bias by subtracting P100 amplitude measured in the V45 condition from the P100 amplitude in the H45 condition. P100 horizontal bias was not significantly different than zero in the left ($M_L = 0.228$, $t_8 = 0.48$, $p = 0.64$, $d = 0.23$) or right ($M_R = 0.285$, $t_8 = 0.49$, $p = 0.64$, $d = 0.23$) hemispheres. We then plotted P100 horizontal bias against full face identification accuracy (Figure 3.4B). The correlation was not significant in either the left ($r = -0.433$, $t_7 = -1.27$, $p = 0.244$) or right ($r = -0.392$, $t_7 = -1.13$, $p = 0.297$) hemispheres. The correlation also was not significant if we averaged P100 horizontal bias scores across hemispheres ($r = -0.420$, $t_7 = -1.23$, $p = 0.260$)

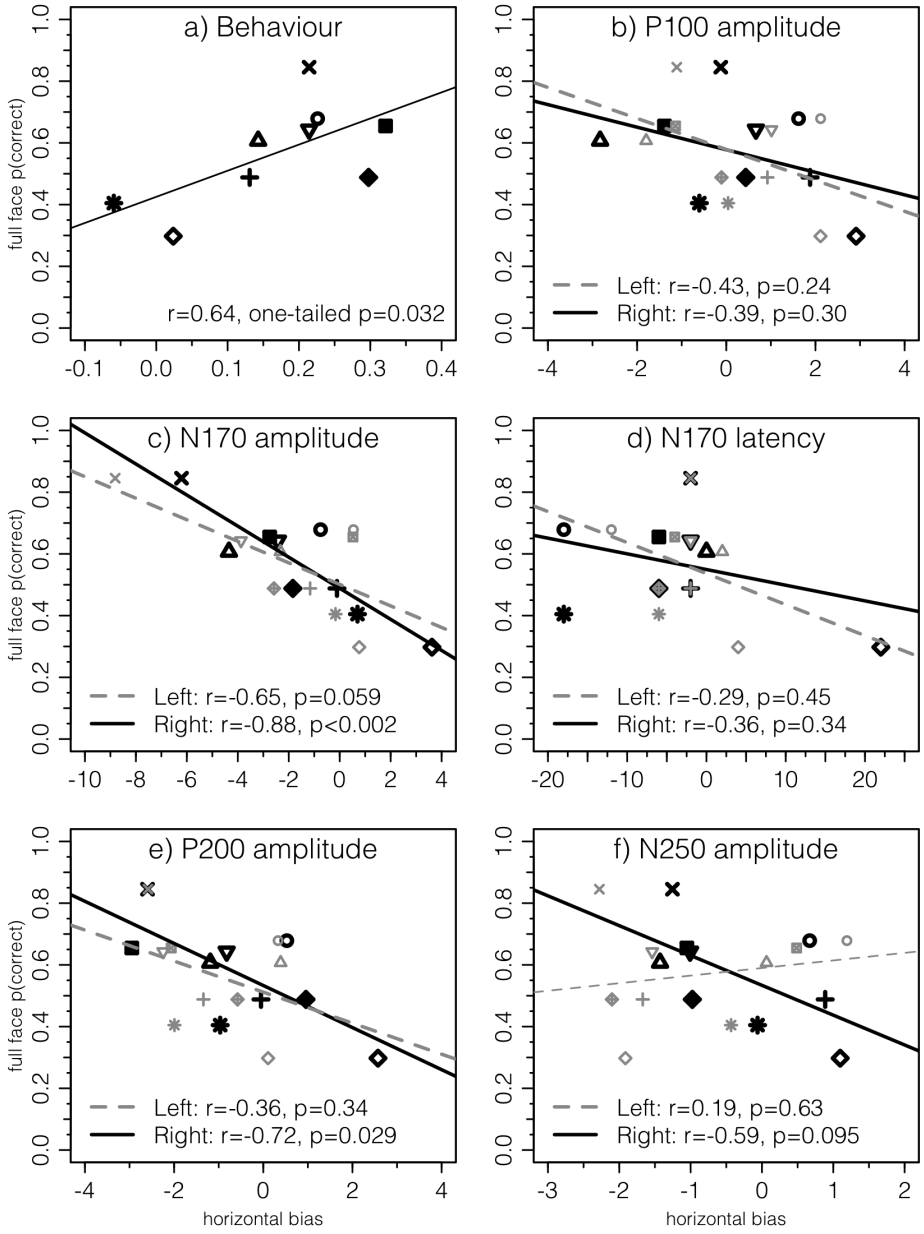

Figure 3.4 – Face identification accuracy (y-axis) plotted as a function of different measures of horizontal bias (x-axis). Horizontal bias was measured by subtracting V45 from H45 using measures of **A)** response accuracy, **B)** N170 latency, **C)** P100 amplitude, **D)** N170 amplitude, **E)** P200 amplitude, and **F)** N250 amplitude. Least squares regression lines are plotted for the left (dashed grey) and right (solid black) hemispheres, and the corresponding Pearson's correlation coefficient is shown in the legend. Single subject symbols match those in Figure 3.3A.

3.4.2.2 N170 Amplitude

N170 amplitude for the right hemisphere is plotted as a function of filter condition in Figure 3.2C. At small filter bandwidths, N170 mean amplitude was smaller in the vertical filter condition than in the horizontal filter condition, but the effect of filter orientation was very small for bandwidths $\geq \pm 72$ deg. The ANOVA revealed that the main effect of *bandwidth* was significant ($F_{(4,32)} = 5.26$, $p_{HF} = 0.0114$, $\eta^2_G = 0.022$), but the main effect of *filter orientation* ($F_{(1,8)} = 2.59$, $p = 0.146$, $\eta^2_G = 0.006$) and the *bandwidth* $\times$ *filter orientation* interaction ($F_{(4,32)} = 1.96$, $p_{HF} = 0.141$, $\eta^2_G = 0.008$) were not significant. The main effect of *hemisphere* ($F_{(1,8)} = 1.28$, $p = 0.290$, $\eta^2_G = 0.010$), as well as its interactions with *filter orientation* ($F_{(1,8)} = 0.021$, $p = 0.889$, $\eta^2_G < 0.001$) and *bandwidth* ($F_{(4,32)} = 0.872$, $p_{HF} = 0.476$, $\eta^2_G < 0.001$) were not significant, and neither was the three-way interaction ($F_{(4,32)} = 0.243$, $p_{HF} = 0.912$, $\eta^2_G < 0.001$).

We also analyzed N170 mean amplitude using an ANCOVA with the P100 mean amplitude as the covariate. As was found in the original ANOVA, N170 amplitude was still subject to a significant main effect of *bandwidth* ($F_{(4,31)} = 5.77$, $p = 0.0014$, $\eta^2_G = 0.025$), and the main effects of *filter orientation* ($F_{(1,7)} = 2.92$, $p = 0.131$, $\eta^2_G = 0.008$) and *hemisphere* ($F_{(1,7)} = 0.143$, $p = 0.717$, $\eta^2_G = 0.001$) were not significant. There also were no significant *filter orientation* $\times$ *bandwidth* ($F_{(4,31)} = 2.14$, $p = 0.100$, $\eta^2_G = 0.008$), *filter orientation* $\times$ *hemisphere* ($F_{(1,7)} = 0.136$, $p = 0.723$, $\eta^2_G < 0.001$), *bandwidth* $\times$ *hemisphere* ($F_{(4,31)} = 1.01$, $p = 0.418$, $\eta^2_G < 0.001$), and *filter orientation* $\times$ *bandwidth* $\times$ *hemisphere* ($F_{(4,31)} = 1.83$, $p = 0.148$, $\eta^2_G < 0.001$) interactions. In summary, including P100 amplitude as a covariate did not appreciably alter the effects of hemisphere, filter orientation, or filter bandwidth on N170 amplitude.

We analyzed the significant main effect of *bandwidth* by computing the linear and quadratic trends of N170 amplitude (averaged across filter orientation) across bandwidth. Both the linear trend ($t_8 = 2.75$, one-tailed $p = 0.0127$, $d = 0.914$) and quadratic trend ($t_8 = 2.34$, one-tailed $p = 0.0238$, $d = 0.779$) were significant. We also compared N170 amplitude at each filter bandwidth to N170 amplitude in the full-face condition, using the Bonferroni adjustment to set familywise α to 0.05 (pairwise $\alpha = 0.01$). Although N170 amplitude was more negative to full faces ($M_{FF} = -5.82$ μV) than to filtered faces at all bandwidths, none of the pairwise comparisons were statistically significant ($M_{\pm 45} = -4.24$, $t_8 = 3.11$, $p = 0.014$, $d = 1.04$; $M_{\pm 54} = -4.90$, $t_8 = 2.03$, $p = 0.076$, $d = 0.678$; $M_{\pm 63} = -5.58$, $t_8 = 0.766$, $p = 0.466$, $d = 0.255$; $M_{\pm 72} = 5.71$, $t_8 = 0.300$, $p = 0.772$, $d = 0.100$; $M_{\pm 81} = -5.78$, $t_8 = 0.131$, $p = 0.899$,

$d = 0.044$).

We calculated N170 amplitude horizontal bias by subtracting N170 amplitude in the V45 condition from the N170 amplitude in the H45 condition. N170 amplitude horizontal bias was not significantly different from zero in the left ($M_L = -1.90$, $t_8 = -1.86$, $p = 0.099$, $d = 0.879$) or right ($M_R = -1.57$, $t_8 = -1.63$, $p = 0.141$, $d = 0.770$) hemispheres. We then correlated N170 amplitude horizontal bias with full face identification accuracy (Figure 3.4C). Full face identification accuracy was significantly correlated with N170 amplitude horizontal bias in the right hemisphere ($r = -0.882$, $t_7 = -4.95$, $p = 0.00174$), and was similar, and in the same direction, in the left hemisphere ($r = -0.649$, $t_7 = -2.26$, $p = 0.0586$). If we averaged N170 amplitude horizontal biases across hemispheres, the correlation was significant ($r = -0.813$, $t_7 = -3.70$, $p = 0.00765$).

3.4.2.3 N170 Latency

N170 latency is plotted in Figure 3.2D. N170 latency for full faces was 200 ms, and inspection of Figure 3.2D indicates that all conditions produced a delayed N170 relative the N170 latency of 177 ms measured in younger adults in Chapter 2. Also, the N170 latency was later for the vertical filter conditions than horizontal filter conditions. The ANOVA revealed a significant main effect of *filter orientation* ($F_{(1,8)} = 7.82$, $p = 0.023$, $\eta^2_G = 0.050$), reflecting an overall later N170 for vertically oriented conditions ($M_V = 206.8$ms) than horizontally oriented conditions ($M_H = 203.8$ ms). Figure 3.2D shows that N170 latency decreased from (approximately) 207 to 204 ms as bandwidth increased from ±45 to ±81 deg; however, the main effect of *bandwidth* ($F_{(4,32)} = 2.68$, $p_{HF} = 0.096$, $\eta^2_G = 0.061$), and the *filter orientation* × *bandwidth* interaction ($F_{(4,32)} = 0.510$, $p_{HF} = 0.585$, $\eta^2_G = 0.013$) were not significant. Despite no significant main effect, the linear trend of *bandwidth* was significant ($t_8 = 1.96$, $p = 0.0428$, $d = 0.654$), and it did not differ between filter orientations ($t_8 = 1.10$, $p = 0302$, $d = 0.360$). The main effect of *hemisphere* ($F_{(1,8)} = 0.026$, $p = 0.876$, $\eta^2_G < 0.001$), and its interactions with *filter orientation* ($F_{(1,8)} = 0.119$, $p = 0.739$, $\eta^2_G < 0.001$) or *bandwidth* ($F_{(4,32)} = 0.494$, $p_{HF} = 0.577$, $\eta^2_G = 0.002$) also were not significant. Finally, the three-way interaction between *filter orientation*, *bandwidth*, and *hemisphere* was not significant ($F_{(4,32)} = 1.15$, $p_{HF} = 0.345$, $\eta^2_G = 0.004$).

We calculated the N170 latency's horizontal bias for each hemisphere. N170 latency horizontal bias was not significantly different than zero for the left ($M_L = -3.11$,

$t_8 = -1.98$, $p = 0.082$, $d = -0.936$) or right ($M_R = -3.56$, $t_8 = -0.909$, $p = 0.390$, $d = -0.428$) hemisphere. We also plotted N170 latency horizontal bias against full face identification accuracy (Figure 3.4D), and found that the correlations were not significant in either the left ($r = -0.287$, $t_7 = -0.794$, $p = 0.453$) or right ($r = -0.363$. $t_7 = -1.03$, $p = 0.337$) hemispheres. N170 latency horizontal bias averaged across hemispheres also was not significantly correlated with full face identification accuracy ($r = -0.351$, $t_7 = -0.991$, $p = 0.355$).

3.4.2.4 P200 amplitude

P200 mean amplitude is plotted as a function of filter bandwidth and orientation in Figure 3.2E. P200 amplitude was smallest for full faces, and was more positive as the filter bandwidth decreased. At most bandwidths, the vertically oriented filter conditions produced a larger P200 than the horizontally oriented filter conditions. The ANOVA revealed a significant main effect of *filter orientation* ($F_{(1,8)} = 10.8$, $p = 0.011$, $\eta^2_G = 0.018$), reflecting a smaller P200 for horizontally filtered ($M_H = -1.05$) than vertically filtered ($M_V = -0.217$) stimuli. The main effect of *bandwidth* ($F_{(4,32)} = 3.50$, $p_{HF} = 0.043$, $\eta^2_G = 0.020$) also was significant, reflecting more positive P200 amplitudes at smaller bandwidths. The *filter orientation* × *bandwidth* interaction ($F_{(4,32)} = 1.33$, $p_{HF} = 0.282$, $\eta^2_G = 0.008$) was not significant. The main effect of *hemisphere* ($F_{(1,8)} = 2.18$, $p = 0.178$, $\eta^2_G = 0.010$) also was not significant, nor were its interactions with *filter orientation* ($F_{(1,8)} = 1.20$, $p = 0.305$, $\eta^2_G < 0.001$) or *bandwidth* ($F_{(4,32)} = 0.659$, $p_{HF} = 0.608$, $\eta^2_G < 0.001$). The three-way interaction ($F_{(4,32)} = 0.257$, $p_{HF} = 0.903$, $\eta^2_G < 0.001$) was not significant.

As indicated in Figure 3.2E, the main effect of *bandwidth* reflects a higher P200 amplitude to smaller filter bandwidths than larger ones. As with the N170 amplitude, we conducted linear and quadratic trend analyses to characterize the effect of bandwidth more accurately. The linear trend of P200 amplitude (averaged across filter orientations) across filter bandwidth was significant ($t_8 = 2.44$, one-tailed $p = 0.020$, $d = 0.812$), but the quadratic trend ($t_8 = 0.842$, $p = 0.212$, $d = 0.281$) was not. Follow up t tests (two-tailed, Bonferroni-corrected $\alpha_{PC} = 0.01$) comparing P200 amplitude (averaged across filter orientations) in each filter bandwidth condition to P200 amplitude in the full face condition ($M_{FF} = -2.10$) revealed that only the difference between the full-face and ±45 bandwidth conditions was significant ($M_{\pm 45} = -0.044$, $t_8 = 3.42$, $p = 0.009$, $d = 1.14$; $M_{\pm 54} = -0.206$, $t_8 = 3.21$, $p = 0.012$, $d = 1.07$; $M_{\pm 63} = -0.727$,

$t_8 = 3.05$, $p = 0.016$, $d = 1.02$; $M_{\pm 72} = -1.07$, $t_8 = 1.96$, $p = 0.085$, $d = 0.654$; $M_{\pm 81} = -1.14$, $t_8 = 2.96$, $p = 0.018$, $d = 0.988$).

P200 amplitude horizontal bias was significantly different than zero in the left hemisphere ($M_L = -1.11$, $t_8 = -2.80$, $p = 0.023$, $d = -1.23$), but not the right hemisphere ($M_R = -0.502$, $t_8 = -0.869$, $p = 0.410$, $d = -0.409$). P200 amplitude horizontal bias is plotted against full face identification accuracy in Figure 3.4E. Full face identification accuracy and P200 horizontal bias were significantly correlated in the right ($r = -0.719$, $t_7 = -2.74$, $p = 0.029$), but not the left ($r = -0.364$, $t_7 = -1.03$, $p = 0.336$) hemisphere. P200 horizontal bias averaged across the two hemispheres was not significantly correlated with full face identification accuracy ($r = -0.626$, $t_7 = -2.12$, $p = 0.0714$).

3.4.2.5 N250 Amplitude

N250 mean amplitude is plotted as a function of filter orientation and bandwidth in Figure 3.2F. As was found with N170 latency and P200 amplitude, N250 amplitude was more negative as bandwidth increased, and also was more negative in the horizontal filter than the vertical filter conditions. The $2 \times 2 \times 2$ ANOVA revealed significant main effects of *filter orientation* ($F_{(1,8)} = 7.91$, $p = 0.023$, $\eta^2_G = 0.08$) and *bandwidth* ($F_{(4,32)} = 5.51$, $p_{HF} = 0.002$, $\eta^2_G = 0.012$), but their interaction ($F_{(4,32)} = 0.905$, $p_{HF} = 0.467$, $\eta^2_G = 0.003$) was not significant. The main effect of *filter orientation* reflected generally more negative N250 amplitudes for horizontally filtered ($M_H = -4.22$) than vertically filtered ($M_V = -3.46$) stimuli, and the main effect of bandwidth reflected generally smaller (less negative) N250 amplitudes for smaller filter bandwidths. The main effect of *hemisphere* ($F_{(1,8)} = 9.89$, $p = 0.014$, $\eta^2_G = 0.013$) was significant, reflecting a slightly more negative amplitude in the left ($M_L = -4.35$) than the right ($M_R = -3.33$) hemisphere. The *filter orientation* $\times$ *hemisphere* interaction ($F_{(1,8)} = 1.02$, $p = 0.341$, $\eta^2_G < 0.001$), *bandwidth* $\times$ *hemisphere* interaction ($F_{(4,32)} = 0.127$, $p_{HF} = 0.972$, $\eta^2_G < 0.001$), and the three-way interaction ($F_{(4,32)} = 0.310$, $p_{HF} = 0.869$, $\eta^2_G < 0.001$) were not significant.

We analyzed the main effect of *bandwidth* by evaluating the linear and quadratic trends of N250 amplitude (averaged across filter orientations) across filter bandwidth. We found that the linear trend was significant ($t_8 = 3.79$, one-tailed $p = 0.003$, $d = 1.26$), but the quadratic trend was not ($t_8 = -0.111$, $p = 0.543$, $d = -0.037$). We also used pairwise comparisons to evaluate differences between N250 amplitude in the

full face condition and N250 amplitude at each filter bandwidth (two-tailed, Bonferonni corrected, $\alpha_{PC} = 0.01$). Although the N250 amplitude to full faces ($M_{FF} = -5.30$) was larger (i.e., more negative) than the N250 in all filter conditions, none of the pairwise differences were statistically significant ($M_{\pm 45} = -3.44$, $t_8 = 2.79$, $p = 0.024$, $d = 0.930$; $M_{\pm 54} = -3.09$, $t_8 = 2.94$, $p = 0.019$, $d = 0.980$; $M_{\pm 63} = -4.12$, $t_8 = 1.97$, $p = 0.084$, $d = 0.658$; $M_{\pm 72} = -4.18$, $t_8 = 1.96$, $p = 0.085$, $d = 0.654$; $M_{\pm 81} = -4.36$, $t_8 = 1.45$, $p = 0.186$, $d = 0.482$).

N250 amplitude horizontal bias is plotted against full face identification accuracy in Figure 3.4F). N250 horizontal bias was not significantly different from zero in either the left ($M_L = -0.911$, $t_8 = -2.16$, $p = 0.063$, $d = -1.02$) or right ($M_R = -0.346$, $t_8 = -1.04$, $p = 0.329$, $d = -0.490$) hemisphere, and was not significantly correlated with full-face identification accuracy in either hemisphere (left hemisphere: $r = 0.189$, $t_7 = 0.510$, $p = 0.626$; right hemisphere: $r = -0.590$, $t_7 = 1.93$, $p = 0.095$), although the correlation in the right hemisphere is similar in magnitude and direction with the correlation obtained with younger adults (Chapter 2). N250 horizontal bias averaged across the two hemispheres was not significantly correlated with full face identification accuracy ($r = -0.215$, $t_7 = -0.583$, $p = 0.578$).

3.4.3 Correlation time-series

Finally, to better characterize the relationship between full face identification accuracy and the horizontal bias of the ERP without being restricted to pre-defined ERP components, we calculated the correlation at every time-point in the ERP separately for the left and right hemispheres (Figure 3.5). We found that the most sustained, relatively strong correlations occurred in the right hemisphere around the N170 peak (186 to 222 ms) and between the P200 and N250 peaks (322 to 382 ms). The sustained correlation between 322 and 382 ms suggests that our earlier analyses of P200 and N250 horizontal bias were likely examining the same underlying neural processing rather than different processing. However, the timing of the correlations, relative to the ERP peaks, is consistent with reports of similar results in younger adults, where the brain-behaviour correlation also occur between the peaks of the P200 and N250 (Figure 2.8 in Chapter 2). The sustained correlation around the N170 peak is novel and not seen in younger adults. There also were correlations at later time points (444-466 ms and 552-572 ms), of which the final correlation corresponds with the N400 peak. A similar result can be seen in younger adults (Chapter 2). For the most part, significant

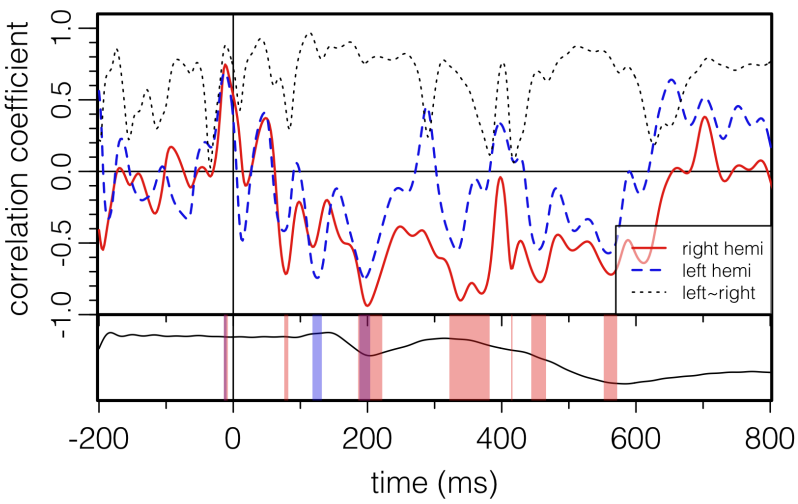

Figure 3.5 – Time-series of the Pearson least squares correlation coefficient (r) between neural horizontal bias (H45 - V45) and full face identification accuracy for the right (red solid) and left (blue dashed) hemispheres. Right and left hemisphere horizontal biases were correlated with each other (black dotted), revealing that horizontal bias was mostly correlated between the two hemispheres. As a temporal reference, the grand-average ERP is shown on the bottom. Shaded red and blue regions reflect statistically significant correlations ($p < 0.05$, uncorrected) for the right and left hemisphere, respectively. Two notable periods of sustained significant correlations occur both occur in the right hemisphere from 186 to 222 ms, and from 322 to 382 ms, which coincide with the peak of N170 and between the P200 and N250, respectively.

correlations were found primarily in the right hemisphere, but they followed a similar pattern in both hemispheres; horizontal bias in the left and right hemispheres were correlated at most times points (dotted black line in Figure 3.5).

3.4.4 Comparison to younger adults

To compare the results here to those obtained in Chapter 2, we combined the data from younger and older adults into a single analysis, separately for each measure. Age group was set as a between-subject factor in mixed-design ANOVAs using type III sums-of-squares. For brevity, we focus primarily on the effects that depend on the factor *age*. The correlations between full face identification accuracy and horizontal bias were compared between younger and older adults using the *cocor* package (Diedenhofen and Musch, 2015), which Fisher-transforms the correlation coefficients and computes a z test on the null hypothesis that the true difference between them is 0. If the difference was not significant, we report a single correlation collapsed across age groups.

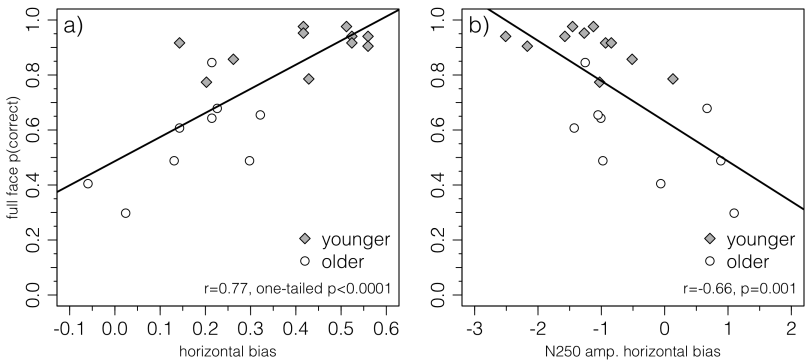

Figure 3.6 – Face identification accuracy (y-axis) plotted as a function of **A)** behavioural horizontal bias and **B)** right hemisphere N250 horizontal bias. Data are plotted for older adults (white) and younger adults (grey, obtained from Chapter 2). As outlined in the text, the correlations were not significantly different between younger and older adults, so Pearson's correlations were calculated using both age groups.

Face identification accuracy was subject to a significant main effect of *age* ($F_{(1,18)} = 54.9$, $p < 0.0001$, $\eta^2_G = 0.679$), reflecting higher accuracy in younger ($M = 0.675$) than older adults ($M = 0.356$). There also was a significant *filter orientation* × *bandwidth* × *age* interaction ($F_{(4,72)} = 15.34$, $p_{HF} < 0.001$, $\eta^2_G = 0.0646$): the *filter orientation* × *bandwidth* interaction was significant in younger ($F_{(4,40)} = 37.2$, $p_{HF} < 0.0001$, $\eta^2_G = 0.254$) but not older ($F_{(4,32)} = 1.08$, $p_{HF} = 0.382$, $\eta^2_G = 0.0091$) adults. The correlation between full face identification accuracy and behavioural horizontal bias was not significantly different between younger and older adults ($\Delta r = 0.175$, $z = 0.469$, $p = 0.639$). When combined across age group, full face identification accuracy and behavioural horizontal bias were significantly correlated ($r = 0.772$, $t_{18} = 5.15$, one-tailed $p = 0.0001$; Figure 3.6A).

P100 amplitude was subject to a significant main effect of *age* ($F_{(1,18)} = 11.1$, $p = 0.0037$, $\eta^2_G = 0.333$), reflecting higher P100 amplitudes in younger ($M = 3.77\mu$V) than older adults ($M = 0.695\mu$V). There also was a *filter orientation* × *hemisphere* × *age* interaction ($F_{(1,18)} = 5.20$, $p = 0.0350$, $\eta^2_G = 0.0008$), reflecting a significant *filter orientation* × *hemisphere* interaction in older ($F_{(1,8)} = 11.1$, $p = 0.0104$, $\eta^2_G = 0.0087$) but not younger adults ($F_{(1,10)} = 1.35$, $p = 0.272$, $\eta^2_G = 0.0004$). No other main effects or interactions were significant. The correlation between full face identification accuracy and P100 horizontal bias was not significantly different between younger and older adults in either the left ($\Delta r = 0.201$, $z = 0.308$, $p = 0.758$) or right hemisphere $\Delta r = 0.149$, $z = 0.308$, $p = 0.758$). When combined across age group, full face

identification accuracy and P100 horizontal bias were not significantly correlated in the left ($r = -0.375$, $t_{18} = -1.71$, $p = 0.104$) or right hemisphere ($r = -0.411$, $t_{18} = -1.92$, $p = 0.0715$). In summary, the effect of filter orientation on P100 amplitude was similar in younger and older adults, with the exception that the effect of filter orientation was right-lateralized in older adults, whereas there was no effect of filter orientation in younger adults.

N170 amplitude was subject to a significant *bandwidth* × *hemisphere* × *age* interaction ($F_{(4,72)} = 3.20$, $p_{HF} = 0.0262$, $\eta^2_G = 0.0009$), reflecting a significant *bandwidth* × *hemisphere* interaction in younger ($F_{(4,40)} = 3.37$, $p_{HF} = 0.0326$, $\eta^2_G = 0.0015$) but not older adults ($F_{(4,32)} = 0.872$, $p_{HF} = 0.491$, $\eta^2_G = 0.0006$). Additionally, the main effects of *filter orientation* and *bandwidth* were significant, but critically, the *filter orientation* × *bandwidth* interaction ($F_{(4,72)} = 6.92$, $p_{HF} = 0.0007$, $\eta^2_G = 0.010$) also was significant, which is unlike the results from the analysis of older adults alone. Furthermore, the *filter orientation* × *bandwidth* × *age* interaction ($F_{(4,72)} = 0.271$, $p_{HF} = 0.831$, $\eta^2_G = 0.0004$) was not significant, indicating that despite the *filter orientation* × *bandwidth* interaction not being significant in older adults, it was not significantly different than the significant *filter orientation* × *bandwidth* in younger adults. The correlation between full face identification accuracy and N170 amplitude horizontal bias was significantly different between younger and older adults in the right ($\Delta r = 0.734$, $z = 2.29$, $p = 0.0222$) but not the left hemisphere ($\Delta r = 0.565$, $z = 1.28$, $p = 0.202$). The significant difference between age groups in the right hemisphere reflected a significant correlation in older adults ($r = -0.882$, $t_7 = -4.95$, $p = 0.0017$) but not in younger adults ($r = -0.148$, $t_9 = -0.449$, $p = 0.664$). In the left hemisphere, we combined across age group and saw that full face identification accuracy and N170 amplitude horizontal bias were not significantly correlated ($r = -0.358$, $t_{18} = -1.63$, $p = 0.121$). In summary, the lack of a significant difference between younger and older adults *filter orientation* × *bandwidth* interaction suggests that the effect of orientation filtering on the N170 amplitude was similar in younger and older adults. However, despite similar effects of orientation filtering, we confirmed that the correlation between N170 amplitude horizontal bias and face identification accuracy was significantly different between younger and older adults.

N170 latency was subject to a significant main effect of *age* ($F_{(1,18)} = 207.3$, $p < 0.0001$, $\eta^2_G = 0.865$), reflecting an earlier N170 for younger ($M = 177\,\text{ms}$) than older adults ($M = 205\,\text{ms}$). The main effect of *filter orientation* was significant, but

it was qualified by a significant *filter orientation* × *age* interaction ($F_{(1,18)} = 7.39$, $p = 0.0141$, $\eta^2_G = 0.0181$), reflecting a significant main effect of *filter orientation* in older ($F_{(1,8)} = 7.82$, $p = 0.0233$, $\eta^2_G = 0.0946$) but not younger adults ($F_{(1,10)} = 0.0199$, $p = 0.891$, $\eta^2_G < 0.0001$). The main effect of *bandwidth* ($F_{(4,72)} = 12.3$, $p_{HF} = 0.0001$, $\eta^2_G = 0.088$) was significant, and its interaction with *age* ($F_{(4,72)} = 0.38$, $p_{HF} = 0.679$, $\eta^2_G = 0.003$) was not significant. All other main effects and interactions were not significant. The correlation between full face identification accuracy and N170 latency horizontal bias was not significantly different between younger and older adults in either the left ($\Delta r = 0.557$, $z = 1.06$, $p = 0.290$) or right hemisphere ($\Delta r = 0.303$, $z = 0.783$, $p = 0.434$). When combined across age group, full face identification accuracy and N170 latency horizontal bias were not significantly correlated in the left ($r = 0.140$, $t_{18} = 0.602$, $p = 0.555$) or right hemisphere ($r = -0.0829$, $t_{18} = -0.353$, $p = 0.728$). In summary, despite the large and significant age-related increase in the N170 latency, N170 latency was affected similarly by filter bandwidth in younger and older adults. However, the effect of filter orientation on N170 latency was greater in older than younger adults.

P200 amplitude was subject to a significant *bandwidth* × *age* interaction ($F_{(4,72)} = 2.80$, $p_{HF} = 0.0493$, $\eta^2_G = 0.0062$), reflecting a significant main effect of *bandwidth* in older ($F_{(4,32)} = 3.50$, $p_{HF} = 0.0435$, $\eta^2_G = 0.0226$) but not younger adults ($F_{(4,40)} = 1.56$, $p_{HF} = 0.203$, $\eta^2_G = 0.0064$). Like the analysis of older adults, the main effects of *filter orientation* and *bandwidth* were significant, and all other main effects and interactions were not significant. The correlation between full face identification accuracy and P200 horizontal bias was not significantly different between younger and older adults in either the left ($\Delta r = 0.0349$, $z = 0.0756$, $p = 0.940$) or right hemisphere ($\Delta r = 0.249$, $z = 0.731$, $p = 0.465$). When combined across age group, full face identification accuracy and P200 horizontal bias were not significantly correlated in the left ($r = -0.004$, $t_{18} = -0.0171$, $p = 0.987$) or right hemisphere ($r = -0.320$, $t_{18} = -1.43$, $p = 0.170$). In summary, the P200 in older adults was more sensitive to orientation filtering than younger adults, but when combined, P200 horizontal bias was still not correlated with face identification accuracy.

N250 amplitude was subject to a significant *bandwidth* × *age* interaction ($F_{(4,72)} = 4.68$, $p_{HF} = 0.0020$, $\eta^2_G = 0.0056$), reflecting a significant main effect of *bandwidth* in older ($F_{(4,32)} = 5.51$, $p_{HF} = 0.0017$, $\eta^2_G = 0.0131$) but not younger adults ($F_{(4,40)} = 2.35$, $p_{HF} = 0.072$, $\eta^2_G = 0.0070$). There also was a significant *hemisphere* × *age* interaction

($F_{(1,18)} = 7.22$, $p = 0.0151$, $\eta^2_G = 0.0146$), reflecting a significant main effect of *hemisphere* for older ($F_{(1,8)} = 9.89$, $p = 0.0137$, $\eta^2_G = 0.0142$) but not younger adults ($F_{(1,10)} = 2.30$, $p = 0.161$, $\eta^2_G = 0.0176$). Like the analysis of older adults, the main effects of *filter orientation* ($F_{(1,18)} = 13.3$, $p = 0.0019$, $eta^2_G = 0.0058$ and *bandwidth* ($F_{(4,72)} = 3.87$, $p = 0.0066$, $eta^2_G = 0.0047$) were significant. The *filter orientation* × *bandwidth* interaction ($F_{(4,72)} = 2.20$, $p_{HF} = 0.0799$, $\eta^2_G = 0.0026$) was not significant, nor was the *filter orientation* × *bandwidth* × *age* interaction ($F_{(4,72)} = 0.82$, $p_{HF} = 0.512$, $\eta^2_G = 0.001$). All other main effects and interactions also were not significant. The failure to find a significant *filter orientation* × *bandwidth* interaction, or the three-way interaction with *age* is surprising because the two-way interaction is significant in younger (Chapter 2) but not older adults. A comparison of Figure 6 in Chapter 2 and Figure 3.2F suggests that the *filter orientation* × *bandwidth* interaction is similar in both age groups, although smaller in older adults; this suggests that the failure to find a significant *filter orientation* × *bandwidth age* interaction is likely due to insufficient power. The correlation between full face identification accuracy and N250 horizontal bias was not significantly different between younger and older adults in either the left ($\Delta r = 0.699$, $z = 1.40$, $p = 0.163$) or right hemisphere ($\Delta r = 0.0362$, $z = 0.0995$, $p = 0.0921$). When combined across age group, full face identification accuracy and N250 horizontal bias were significantly correlated in the right ($r = -0.665$, $t_{18} = -3.77$, $p = 0.0014$; Figure 3.6B), but not the left hemisphere ($r = 0.0006$, $t_{18} = 0.003$, $p = 0.998$). In summary, N250 amplitudes of younger and older adults were similarly affected by filter bandwidth and orientation. Unlike the analysis described in Chapter 2, the current, combined analysis suggests that the effect of bandwidth was not modulated significantly by filter orientation in either age group.

In summary, these analyses found that full face identification accuracy was correlated significantly with horizontal bias measured using behaviour, right hemisphere N170 amplitude (older adults only), and right hemisphere N250 amplitude. Next, we assessed if these measures were related by correlating them with each other. Note that we expect these measures to be correlated to some degree because they are all based on responses to the same stimuli on the same trials, unlike the correlations with full face identification accuracy, which were based on an independent set of trials. Indeed, the correlations between all three indices of horizontal bias were significant (behaviour & N170: $r = -0.510$, $t_{18} = -2.51$, $p = 0.0217$; behaviour & N250 $r = -0.624$, $t_{18} = -3.39$, $p = 0.0033$; N170 & N250 $r = 0.782$, $t_{18} = 5.33$, $p < 0.0001$).

Next, we used analysis of covariance (ANCOVA) to assess if age differences in face identification accuracy were related to age differences in horizontal bias. Without any covariate, face identification accuracy differed significantly between the two age groups ($F_{(1,18)} = 38.0$, $p < 0.0001$, $\eta_p^2 = 0.678$). When behavioural horizontal bias was the covariate ($F_{(1,17)} = 42.2$, $p < 0.0001$, $\eta_p^2 = 0.713$), the main effect of *age* ($F_{(1,17)} = 11.7$, $p = 0.0033$, $\eta_p^2 = 0.407$) was $\approx 40\%$ smaller but still significant. When right hemisphere N170 amplitude horizontal bias was the covariate ($F_{(1,17)} = 23.1$, $p = 0.0002$, $\eta_p^2 = 0.576$), the main effect of *age* ($F_{(1,17)} = 55.3$, $p < 0.0001$, $\eta_p^2 = 0.765$) actually increased slightly and was still significant. Finally, when right hemisphere N250 horizontal bias was the covariate ($F_{(1,17)} = 34.2$, $p < 0.0001$, $\eta_p^2 = 0.668$), the main effect of *age* ($F_{(1,17)} = 26.2$, $p < 0.0001$, $\eta_p^2 = 0.607$) was approximately the same and remained significant. In all ANCOVAs, the effect of the covariate was significant and accounted for over 57% the variance in each case.

The ANCOVA analyses indicate that the age difference in face identification accuracy is not entirely accounted for by changes in any single measure of horizontal bias. However, it is possible that the measures *collectively* account for the age difference. To test this idea, we fit the face identification data with a linear model that included all three measures of horizontal bias as covariates and found that the main effect of age was still significant ($F_{(1,15)} = 15.4$, $p = 0.0014$, $\eta_p^2 = 0.506$). Finally, to examine whether the three measures of horizontal bias accounted for the same or different portions of variation in identification accuracy, we examined the effects of varying the order in which the three covariates were entered into the linear model. The results are summarized in Table 3.1. When the first covariate was behavioural horizontal bias (Models A & B), then the variation accounted for by both of the N170 and N250 biases was drastically reduced; however, when behavioural horizontal bias was the final covariate (Models C & D), it still accounted for significant variation in identification accuracy. Interestingly, when N170 amplitude horizontal bias was the first covariate, the N250 horizontal bias was still a significant covariate; however, when N250 horizontal bias was the first covariate, N170 amplitude horizontal bias was no longer a significant covariate. These results suggest that i) the behavioural measure of horizontal bias accounted for variation that was not accounted for by either ERP measure; and ii) that the N250 measure of bias accounted for variation that was not accounted for by the N170 measure; and iii) the N170 measure did not account for any unique variation. Finally, the ANCOVA analyses suggest that other factors besides horizontal

bias contribute to age differences in face identification.

Model A:	F	p	η_p^2
Behaviour	52.4	<0.0001	0.777
N170 amp	1.16	0.298	0.072
N250	4.03	0.063	0.212
Age	15.4	0.0014	0.506
Model B:			
Behaviour	52.4	<0.0001	0.777
N250	4.82	0.044	0.243
N170 amp	0.368	0.553	0.024
Age	15.4	0.0014	0.506
Model C:			
N170 amp	21.3	0.0003	0.587
N250	17.7	0.0008	0.542
Behaviour	18.5	0.0006	0.553
Age	15.4	0.0014	0.506
Model D:			
N250	38.8	<0.0001	0.721
N170 amp	0.175	0.681	0.012
Behaviour	18.5	0.0006	0.553
Age	15.4	0.0014	0.506

Table 3.1 – Results form four ANCOVA models that included the three covariates (i.e., measures of horizontal bias based on behaviour, N170 amplitude, and N250 amplitude) and one between-subjects factor (age). The models differed in terms of the order in which the covariates were entered into model. The table shows the F and p values for the Type I (sequential) sums-of-squares, and a measure of association strength (η_p^2), for each term in the model.

3.5 Discussion

We measured response accuracy and ERPs while in older adults who performed a 1-of-6 face identification task that used faces that contained varying amounts of horizontal or vertical structure. We then correlated measures of behavioural and neural *horizontal bias* with full face identification accuracy. In comparison to younger adults who completed the same experiment (Chapter 2), and consistent with previous studies of face perception in older adults (Grady et al., 1994; Konar et al., 2013; Obermeyer

et al., 2012), we found that older adults made more errors in a face identification task and had a weaker horizontal bias: not only was response accuracy worse overall, differences in response accuracy to horizontally and vertically filtered stimuli was smaller in older than younger adults. Still, we found that behavioural horizontal bias was significantly correlated with full face identification accuracy. The correlation ($r = 0.637$) did not differ significantly from the value ($r = 0.46$) obtained with younger adults in Chapter 2 or by Pachai et al. (2013, $r = 0.52$). Similar to the results obtained with response accuracy, we found that older adults exhibited ERPs to horizontally and vertically filtered faces that were less differentiated (Figure 3.2C-F) than ERPs in younger adults (Chapter 2). The observed correlations between horizontal bias and identification accuracy is yet another indication that identification is associated with the preferential processing of horizontally oriented facial structure (Dakin and Watt, 2009; Goffaux and Dakin, 2010; Pachai et al., 2013), and that between-group differences in face identification accuracy are due, at least partly, to differences in horizontal bias. For example, other studies have shown horizontal bias is weaker in persons with prosopagnosia (Pachai et al., 2015), and is stronger for familiar than unfamiliar faces in healthy adults (Pachai et al., 2017).

The N170 of older adults showed both similarities and differences to that of younger adults. Most obviously, we replicated previous reports that the N170 in older adults was delayed relative to the N170 in younger adults (Wiese et al., 2008; Rousselet et al., 2009). However, we also found that the effect of orientation of facial structure on the N170 was similar to that found in younger adults (Chapter 2): in both age groups, N170 amplitude increased when horizontal structure was removed, whereas N170 latency increased when horizontal or vertical structure was removed. The results are consistent with the idea that N170 amplitude is sensitive to facial information about the eyes (Bentin et al., 1996; Itier et al., 2007; Rousselet et al., 2014), whereas N170 latency is sensitive to accrual of all face-related information (Rousselet et al., 2009, 2010; Bieniek et al., 2013). This interpretation of the effect of filter orientation on N170 latency is consistent with the observation that vertical structure carries less information about a face's identity than horizontal (Pachai et al., 2013, 2018), which should result in a slower rate of information accrual for vertical structure. Additionally, the greater sensitivity of human observers to horizontal facial structure (Pachai et al., 2013, 2018) would further exacerbate differences in the rate of information accrual. Our experiments cannot dissociate the potential influences of differential stimulus information and

perceptual sensitivity on the effects of filter orientation on the N170 latency. It may be possible to distinguish the effects of these two factors using stimuli that contain normal facial structure at all orientations but which are diagnostic structure only at particular orientations (see stimuli in Pachai et al., 2018, or in Chapter 4). This method would ensure that the visual system receives a input of facial structure at all orientations, but task-relevant information at only certain orientation bands. Therefore, behaviour would be driven by the perceptual sensitivity to the task-relevant structure than to low level processing of stimulus information.

Face identification relies, in part, on the efficient use of horizontal facial structure, so our result showing a significant correlation between N250 horizontal bias and full face identification accuracy is consistent with the proposition that the N250 is sensitive to the identity of a face (Tanaka et al., 2006; Kaufmann et al., 2009). Younger adults' full face identification accuracy is correlated with their N250 horizontal bias (Chapter 2), and we see a similar relationship in older adults, albeit weaker and right-lateralized (Figure 3.4F). Importantly, full face identification accuracy in older adults, but not younger adults, was significantly correlated with the horizontal bias of the N170 amplitude (Figure 3.4C). This correlation may be associated with low-level visual constraints on face perception in older adults (Grady et al., 1994; Owsley et al., 1981). According to this idea, face identification accuracy in younger and older adults may be limited primarily by the strength of their horizontal bias, which is reflected in the N250, but only older adults are significantly constrained by the amount of task-relevant information encoded by the early stages of visual processing, which is reflected in the N170. That is to say, older adults may still have a bias towards processing horizontal structure, but their behaviour is affected by the effects of horizontal bias on behaviour as well as the effects of the fidelity of signal in the visual system. The ANCOVA results presented in Table 3.1 are relevant here. Those analyses showed that the association between face identification accuracy and the measures of horizontal bias based on N170 and N250 amplitudes depended on the order in which those terms were entered into the linear model. In particular, a comparison of Models C & D shows that the N170 term accounts for significant variation in accuracy when it appears before, but not after, the N250 term, but the N250 term accounts for significant variation when it appears before or after the N170 term. This order dependency indicates that the N250 term accounts for most of the N170-related variance plus a unique component of variance, and it is consistent with the notion that processes that affect the N170

measure of horizontal bias are partly responsible for affecting the N250 measure of horizontal bias. In this framework, the fact that the N170 amplitude measure of horizontal bias was correlated with full face identification accuracy only in older adults means that age-related decreases in the amount of information encoded by early visual processes alter the N170 (Rousselet et al., 2010; Bieniek et al., 2013) and place greater constraints on face identification in older adults. Of course, these claims, which are based on post-hoc analyses performed on a fairly small data set, are preliminary and should be evaluated in future experiments using a larger sample size and planned comparisons.

In conclusion, the experiment in this chapter showed that face perception in older adults is significantly constrained by horizontal face information. Relative to younger adults, older adults have a weaker bias to processing the horizontal structure of a face, which may be due to a combination of changes in low level visual processing and changes in the perceptual process(es) tuned to horizontally oriented facial structure. Although much of the age-related deficits in our sample were accounted for by decreases in behavioural, N170 amplitude, and N250 horizontal biases, there still remained age-related changes not accounted for by our measures of horizontal bias. Future studies should investigate how changes in face identification accuracy correspond to changes in horizontal bias to link recovery from age-related deficits to the strengthening of horizontal bias. As our interpretation of the N170 results suggests, care should be taken to minimize the effects of low level stimulus differences on early neural processes, and instead focus on changes in the perceptual sensitivity to horizontal information.

References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–384.

Balas, B. J., Schmidt, J., and Saville, A. (2015). A face detection bias for horizontal orientations develops in middle childhood. *Frontiers in Psychology*, 06(June):1–8.

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, 8(6):551–565.

Bieniek, M. M., Frei, L. S., and Rousselet, G. A. (2013). Early ERPs to faces: aging, luminance, and individual differences. *Frontiers in Psychology*, 4(May):268.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–6.

Chaby, L., Narme, P., and George, N. (2011). Older adults' configural processing of faces: Role of second-order information. *Psychology and Aging*, 26(1):71–79.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1):2–5.

Dakin, S. C. and Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., and Lueschow, A. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *NeuroImage*, 35(4):1495–1501.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21.

Diedenhofen, B. and Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10(4):1–12.

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

Folstein, M. F., Robins, L. N., and Helzer, J. E. (1983). The mini-mental state examination. *Archives of general psychiatry*, 40(7):812–812.

Gaspar, C. M., Bennett, P. J., and Sekuler, A. B. (2008). The effects of face inversion and contrast-reversal on efficiency and internal noise. *Vision Research*, 48:1084–1095.

Goffaux, V. and Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, 1(September):1–14.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., and Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, 81:1–11.

Goffaux, V. and Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, 6:34204.

Goffaux, V., Poncin, A., and Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 Year Old). *PLoS ONE*, 10(9):1–17.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39:3537–3560.

Grady, C. L., Ma Maisog, J., Horwitz, B., Ungerleider, L. G., Mentis, M. J., Salerno, J. A., Pietrini, P., Wagner, E., Haxby, J. V., Gillette, J., Giacometti, K., Baldwin, P., Jacobs, G., Stein, S., Green, S., Fluck, S., and Der, M. (1994). Age-related changes in cortical blood flow activation during visual processing of faces and location. *The Journal of Neuroscience*, 14(3):1450–1462.

Itier, R. J., Alain, C., Sedore, K., and McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience*, 19(11):1815–1826.

Jacques, C., Schiltz, C., and Goffaux, V. (2014). Face perception is tuned to horizontal orientation in the N170 time window. *Journal of Vision*, 14:1–18.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11):4302–4311.

Kaufmann, J. M., Schweinberger, S. R., and Burton, A. M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience*, 21(4):625–41.

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., Broussard, C., and Others (2007). What's new in Psychtoolbox-3. *Perception*, 36(14):1.

Konar, Y., Bennett, P. J., and Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, 21(1):38–43.

Konar, Y., Bennett, P. J., and Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision Research*, 88:38–46.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(NOV):1–12.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*, volume 102. The MIT Press.

Maurer, D., Le Grand, R., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6):255–260.

Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2):61–64.

Murphy, J. and Cook, R. (2017). Revealing the mechanisms of human face perception using dynamic apertures. *Cognition*, 169:25–35.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., and Chertkow, H. (2005). The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699.

Obermeyer, S., Kolling, T., Schaich, A., and Knopf, M. (2012). Differences between old and young adults' ability to recognize human faces underlie processing of horizontal information. *Frontiers in aging neuroscience*, 4(April):3.

Olejnik, S. and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, 8(4):434–447.

Owsley, C., Sekuler, R., and Boldt, C. (1981). Aging an low-contrast vision: face perception. *Investigative Ophthalmology & Visual Science*, 21(2):362–365.

Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2018). The bandwidth of diagnostic horizontal structure for face identification. *Perception*, January 19:1–17.

Pachai, M. V., Bennett, P. J., Sekuler, A. B., Corrow, S., and Barton, J. J. S. (2015). Sensitivity to horizontal structure and face identification in developmental prosopagnosia and healthy aging. *Canadian Journal of Experimental Psychology*, 69(4):341.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to Information Conveyed by Horizontal Contours is Correlated with Face Identification Accuracy. *Frontiers in Psychology*, 4(74).

Pachai, M. V., Sekuler, A. B., Bennett, P. J., Schyns, P. G., and Ramon, M. (2017). Personal familiarity enhances sensitivity to horizontal structure during processing of face identity. *Journal of Vision*, 17(6):5.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics Transforming numbers into movies. *Spatial Vision*.

Piepers, D. W. and Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, 3(DEC):1–11.

R Core Team and R Development Core Team, R. (2017). R: A Language and Environment for Statistical Computing.

Riesenhuber, M., Jarudi, I., Gilad, S., and Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(Suppl 6):S448–S450.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., and Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11(1):69–74.

Rousselet, G. A., Gaspar, C. M., Pernet, C. R., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2010). Healthy aging delays scalp EEG sensitivity to noise in a face discrimination task. *Frontiers in Psychology*, 1(JUL):1–14.

Rousselet, G. A., Husk, J. S., Pernet, C. R., Gaspar, C. M., Bennett, P. J., and Sekuler, A. B. (2009). Age-related delay in information accrual for faces: evidence from a parametric, single-trial EEG approach. *BMC Neuroscience*, 10:114.

Rousselet, G. A., Ince, R. A. A., van Rijsbergen, N. J., and Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *Journal of Vision*, 14(13):1–24.

Rousselet, G. A., Macé, M., and Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision*, 4:13–21.

Sadeh, B., Podlipsky, I., Zhdanov, A., and Yovel, G. (2010). Event-related potential and functional MRI measures of face-selectivity are highly correlated: A simultaneous ERP-fMRI investigation. *Human Brain Mapping*, 31(10):1490–1501.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–6.

Tanaka, J. W., Curran, T., Porterfield, A. L., and Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, 18(9):1488–97.

Tanaka, J. W. and Farah, M. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46A(2):225–45.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87:154–163.

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

Wang, R., Li, J., Fang, H., Tian, M., and Liu, J. (2012). Individual Differences in Holistic Processing Predict Face Recognition Ability. *Psychological Science*, 23(2):169–177.

Wiese, H., Schweinberger, S. R., and Hansen, K. (2008). The age of the beholder: ERP evidence of an own-age bias in face memory. *Neuropsychologia*, 46(12):2973–2985.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., Bub, D. N., and Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance*, 36(1):122–135.

Williams, C. and Henderson, J. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition*, 35(8):1977–1985.

Chapter 4

Practice identifying faces recovers horizontal bias in older adults

4.1 Abstract

Face identification relies on increased sensitivity to horizontal relative to vertical facial structure (i.e., horizontal bias). Face identification accuracy declines with age, and so does horizontal bias. Declines in face identification are associated with declines in horizontal bias. Additionally, face identification can improve with perceptual learning. Here, we asked if practice identifying faces leads to changes in the horizontal bias of older adults. Older adult observers (M = 70.8 years) were trained in a 1-of-10 face identification task with unfiltered faces. Before and after training, we recorded response accuracy to faces filtered to contain informative structure in certain orientation bands, and non-informative facial structure in orthogonal orientation bands. The resulting stimuli appeared like intact faces, but had informative structure in orientations bands that were unspecified to the observer. Horizontal bias at pre-training was negligible and not correlated with overall face identification accuracy, but horizontal bias at post-training was significant and positively correlated with full face identification accuracy. Results suggest that older adults deficits in face identification may be linked to a weakened horizontal bias, but that improved horizontal bias underlies improvements seen in training.

4.2 Introduction

For most adults, face discrimination and identification appear to be effortless tasks. Our ability to perceive upright faces rapidly and accurately is likely due to the fact that we regularly interact with faces in everyday life (Valentine, 1988). Upright face expertise contrasts with the relative difficulty in identifying inverted faces (Yin, 1969), and this so-called face inversion effect is consistent with studies of perceptual learning in which training to identify textures produces texture inversion effects (Hussain et al., 2009a) which are long lasting (Hussain et al., 2011). In fact, training to identify inverted faces abolishes the FIE (Hussain et al., 2009b), further supporting the notion that better identification of upright faces is due to experience with upright faces.

What causes the face inversion effect? One possibility is that upright and inverted faces are processed with similar mechanisms, but lifelong learning has resulted in those mechanisms operating more efficiently with upright faces (Sekuler et al., 2004; Riesenhuber et al., 2004; Williams and Henderson, 2007; Gaspar et al., 2008; Willenbockel et al., 2010; Pachai et al., 2013). This idea is supported by studies showing that horizontal facial structure, which conveys a great deal of information about a face's identity, is processed more efficiently in upright than inverted faces (Pachai et al., 2013; Dakin and Watt, 2009). Indeed, face identification accuracy relies on intact horizontal structure (Dakin and Watt, 2009), and there are individual differences in how sensitive an individual is to horizontal relative to vertical face structure (Pachai et al., 2013). Differential sensitivity to horizontal and vertical facial structure, referred to as *horziontal bias*, varies across individuals, and this variation is correlated with face identification accuracy (Pachai et al., 2013; Pachai et al., 2018; Chapters 2 & 3), and with individual differences in the magnitude of the face inversion effect (Pachai et al., 2013). Horizontal bias also appears to be affected by visual experience with faces. For example, horizontal bias increases through childhood (Balas et al., 2015; Goffaux et al., 2015), and, in adults, is stronger for familiar than unfamiliar faces (Pachai et al., 2017). The effect of familiarity on horizontal bias is further evidence that the process(es) underlying face perception is/are experience-dependent.

Healthy ageing is associated with a decline in the ability to identify faces (Grady et al., 1994; Konar et al., 2013) and a reduction in horizontal bias (Obermeyer et al., 2012; Goffaux et al., 2015). Results presented in Chapter 3 show that although older adults exhibit a reduced horizontal bias, face identification accuracy and horizontal

bias are correlated with each other. The experiment in Chapter 3 showed that, as is the case in younger adults (Chapter 2), face identification accuracy in older adults was correlated with the horizontal bias of the N250, an event-related potential (ERP) component known to be sensitive to face identity information (Tanaka et al., 2006; Kaufmann et al., 2009). However, unlike younger adults, identification accuracy in older adults also was correlated with the horizontal bias of the N170, an ERP component that is affected by low-level stimulus information (Rousselet et al., 2008), including the orientation of the facial structure (Chapter 2). We suggested that the correlation between identification accuracy and the N170 in older but not younger adults is caused by early visual processes placing greater constraints on face identification accuracy in older adults than younger adults.

One way to assess horizontal bias is by testing identification accuracy with face stimuli filtered to contain structure only at certain orientation bands (e.g., Chapters 2 and 3). Although this method is sensitive to the differential perceptual sensitivity to horizontal and vertical facial structure, it may also be affected by the changes in the amount of task-relevant information. Horizontal structure typically contains more information about face identity than vertical structure (Pachai et al., 2013, 2018). Therefore, even without a bias to process horizontal structure, identification accuracy should be better for faces containing only horizontal structure than only vertical structure (see Figure 4.1 B & C for a simplified model). This difference in diagnostic information is an especially important issue if face perception in older adults is disproportionally affected by early visual processes. For example, consider the possibility that early visual deficits in older adults may prevent them from discovering and/or extracting the diagnostic horizontal structure efficiently from an intact (unfiltered) face, but once the information is past early vision processes, it is processed just as well as younger adults. If the stimulus contains only horizontal structure, then older adults need not discover and extract the diagnostic information from a full face, since the to-be-extracted information is presented in isolation. With that in mind, it is possible that our previous findings showing associations between face identification accuracy and the N170 may have underestimated the effects of early processes, and subsequently underestimated age differences in the ERP horizontal bias. Hence, one goal of the current experiment is to evaluate horizontal bias in older adults using stimuli from Pachai et al. (2018) in which observers must discover the diagnostic information from within an entire face. Specifically, instead of just removing

oriented structure as we did in Chapter 3, the oriented structure was replaced with uninformative facial structure (see Figures 4.1 D & E & 4.2). For example, in a 1-of-10 face identification task, a stimulus may contain diagnostic structure from a target face in a narrow horizontal orientation band, and non-diagnostic structure from the average of all 10 faces in all remaining orientation bands. Together, the stimulus resembles an intact face, yet only a particular orientation band contains diagnostic structure. A more face-like stimulus allows for the measuring of how identification accuracy is affected by differently informative structure in different orientation bands, while increasing the likelihood that observers use their usual face processing strategy in all stimulus conditions. Figure 4.1 depicts a simplified model where the strategy is to assign higher weights to information in horizontally tuned orientation channels than vertically tuned orientation channels. Figures 4.1 D & E demonstrate how such a system will behave when the stimulus contains structure at all orientations, but the structure is informative only in the horizontal (D) or only in vertical (E) orientation bands.

Another goal of this experiment is to see if face perception in older adults can improve with practice and, critically, if improvements are associated with changes in horizontal bias. Horizontal bias is greater for familiar faces relative to unfamiliar faces (Pachai et al., 2017), and therefore we expect to see that practice-induced improvements in face identification are associated with increases in horizontal bias for those faces. In the current experiment, older adults were trained in a 1-of-10 identification task with a set of 10 unfiltered faces for three consecutive days. We measured their behavioural horizontal bias with filtered faces before and after training, and correlated those measures of horizontal bias with identification accuracy for unfiltered faces.

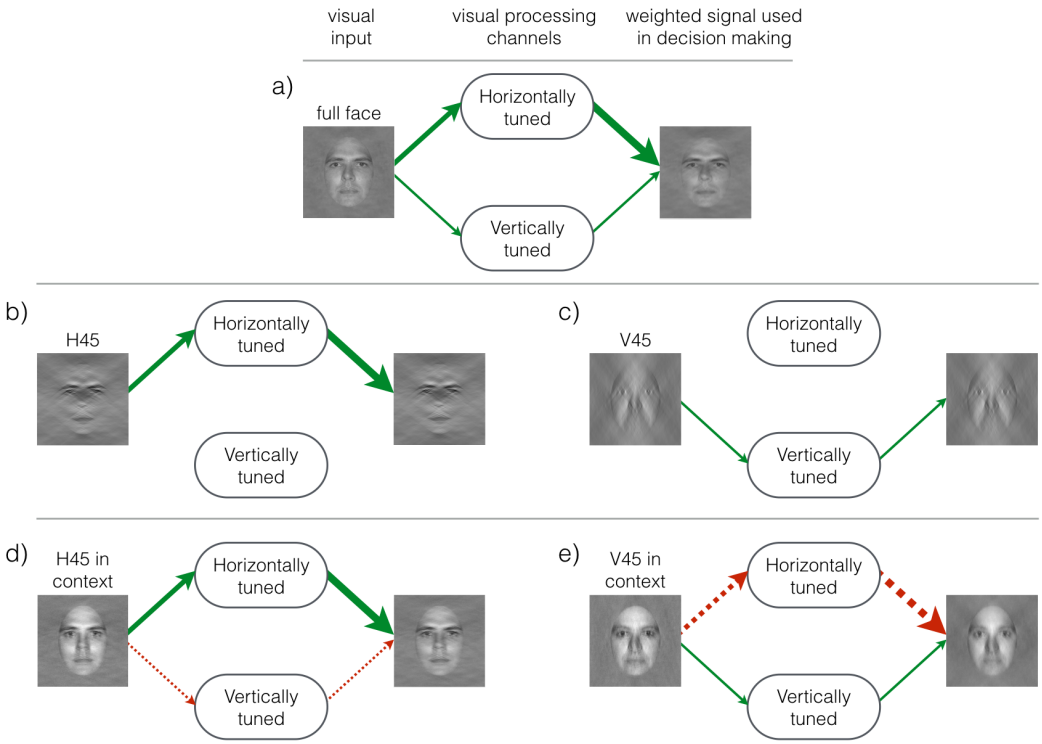

Figure 4.1 – An illustration of the effects of re-weighting horizontal and vertical facial structure to reflect differences in the amount of information conveyed by the two bands of orientations. In each panel, the faces on the left represent the input to the visual system, and the faces on the right represent task-relevant information at the level of the decision variable. The arrows on the left represent the signal entering the visual system, and the arrows on the right represent the processed signal reaching the decision variable. Green and red arrows indicate that the orientation band does or does not contain task-relevant information, respectively. Thickness of arrows/lines reflect the amount of facial information carried in/by a channel. For example, faces intrinsically have more facial information in horizontal than vertical orientation structures, hence the slightly thicker arrow for the horizontal signal on the left. Efficient processing would place greater weight on horizontal structure, hence an even thicker line/arrow for the horizontal signal on the right. **A)** Structure at all orientations is visible and informative. **B)** Horizontal structure is the only structure visible, hence no signal in the vertically tuned input and processing channels. **C)** Vertical structure is the only structure visible, hence no signal in the horizontally tuned input and processing channels. **D)** Structure is visible at all orientations, but horizontal structure is informative whereas vertical structure is not. **E)** Reverse of D. Notice that after horizontally biased processing, the signal at the decision making level better reflects the target identity in D, but is less representative of the target identity in E. We assumed that re-weighting occurs only for the horizontal signal (up-weighted) and that the weight of the vertical channels remains true to the incoming signal. If we assumed that vertical information is down-weighted, the decision-level estimates would veer to an even better representation of the target in D, and a worse representation in E.

4.3 Methods

4.3.1 Subjects

Twelve older adults (6 male; range = 67-77 years old, M=70.8, SD=3.55) from the surrounding community participated in the experiment. All participants had normal or corrected-to-normal Snellen acuity, a normal Pelli-Robson contrast sensitivity (range=1.65-1.95, M=1.86, SD=0.14), and normal MoCA (range=24-30, M=27.5, SD=1.88; Nasreddine et al., 2005) and MMSE scores (range=27-30, M=28.5, SD=1.17; Folstein et al., 1975).

4.3.2 Apparatus & Stimuli

Stimuli were generated on an Apple Macintosh G5 computer using MATLAB 7.10.0 (Mathworks, 2007) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Stimuli were presented on a 17-inch NEC MultiSync FE700+ CRT monitor with a resolution of 1024 × 768 (32 pixels/cm) and an 85 Hz refresh rate. Viewing was binocular through natural pupils from a viewing distance of 100 cm. The monitor had an average luminance of 49.7 cd/m^2. Stimuli were centered on a 256 × 256 pixel matrix and subtended 4.6 deg of visual angle. Stimuli were generated using the 10 faces from Gold et al. (1999a), and were each filtered with an ideal band-pass orientation filter centered on either 0 deg (horizontal) or 90 deg (vertical). Filter bandwidth varied form ±10 deg to ±70 deg in steps of ±10 deg, resulting in 16 filter conditions (Figure 4.2A): eight different filter bandwidths centered on horizontal, and eight different filter bandwidths centered on vertical. In a seventeenth condition, the filter bandwidth was ±90 deg, which covers all orientations: we refer to this as the 'full face' condition. For the filtered conditions, the empty orientation bands were filled with non-informative structure, which we refer to as a face context. The face context was derived by computing a pixel-by-pixel average of all 10 faces and therefore was non-diagnostic in an identification task. In each filtered-face condition, the orientations that were removed by the filtering operation were replaced by structure at the corresponding orientations in the face context. Prior to filtering, target and context each had an RMS contrast of 0.2. After filtering and combining the two, the final stimulus RMS contrast was fixed at 0.2. The stimulus was embedded in a white noise that had an RMS contrast of 0.1.

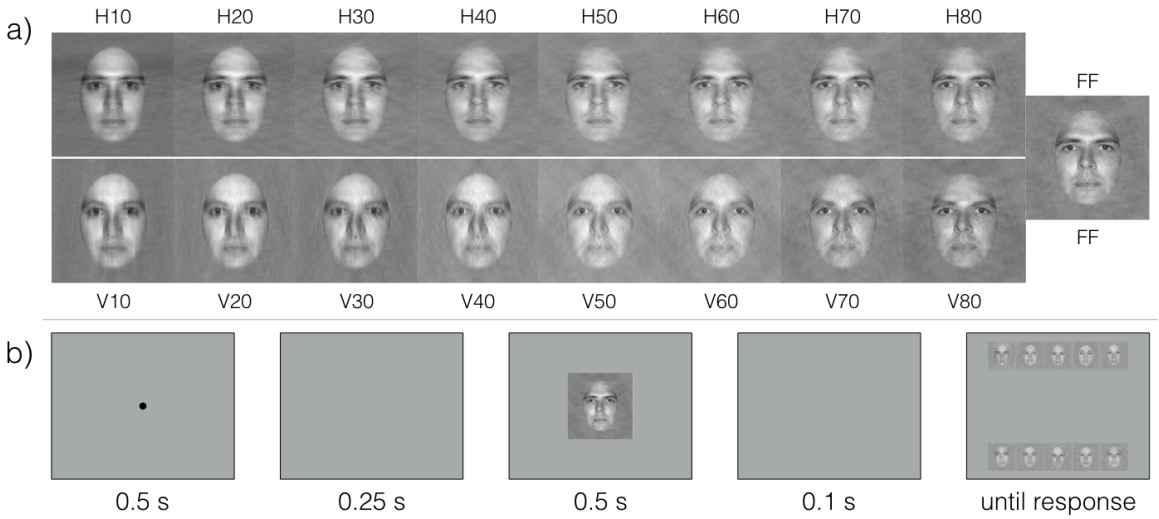

Figure 4.2 – **A)** Example stimuli of the 17 different conditions using in the pre- and post-training sessions. Only the full face condition was used during training trials. **B)** Visualization of the trial structure. During both training and testing, response screen alternatives were full faces.

4.3.3 Procedure

The trial structure is pictured in Figure 4.2B. Each trial began with a central black fixation point on a grey screen for 0.5 s, followed by a blank screen for 0.25 s. The stimulus was then presented for 0.5 s, followed by a 0.1 s blank screen, and then, finally, the response screen with the 10 alternatives until response selection. The ten response screen thumbnails each subtended 2.6 deg of visual angle, had 0.4 RMS contrast, always were full faces without any filtering, and were presented in the same spatial arrangement on every trial. Participants indicated their response by clicking on one of the faces with a computer mouse. Auditory feedback was provided after every response in the form of 600 and 200 Hz tones after correct and incorrect trials, respectively. The next trial started immediately following the response.

4.3.4 Design

During training, the stimulus was always a full face, and face identities were randomly interleaved. Each face identity was presented on 120 trials, for a total of 1200 trials of training. Training trails were separated across three consecutive days (400 trials per session), with exactly 24 hours between sessions. Participants completed testing trials

on the days immediately before and after training. Test trials were identical to training trials, with the exception that the stimuli included all sixteen filtered conditions and the full-face condition. Each condition was repeated 20 times (twice per identity) for a total of 340 trials (10 identities × 17 conditions × 2 repetitions per condition) in each testing session. In total, participants completed 5 consecutive sessions: pre- and post-training test sessions, and three training sessions. At the start of each session, participants completed 10 practice trials with unfiltered faces.

4.4 Results

To maintain an orthogonal design in the analysis of testing data, the full face condition was omitted from the analyses of variance (ANOVA) on testing data. For any effect including the factor *bandwidth*, we did not assume sphericity and report Huynh-Feldt corrected p-values instead. Effect size was measured using Cohen's d for t tests, and generalized eta squared (η^2_G) for repeated-measure ANOVAs, as described by Bakeman (2005) and Olejnik and Algina (2003). Lakens (2013) suggests that Cohen (1988)'s suggested benchmarks of effect size can be used when interpreting η^2_G (small: 0.01, medium: 0.04, large: 0.16). For t tests, we report effect size using Cohen's d. Statistical analyses were performed with R (R Core Team and R Development Core Team, 2017).

Response accuracy during the training sessions was calculated for 12 blocks of 100 trials. Results from the training sessions are displayed in Figure 4.3. Overall, participants response accuracy increased from 65% to 89%. One participant performed at ceiling throughout all of training, and another reached ceiling within the first session. A third participant did not complete all trials during the first training session, but did complete all of the trials in sessions 2 and 3. Otherwise, all participants showed an increasing trend across training bins. Training data were analyzed using linear trend analysis. Linear trends were calculated for each session, and submitted to a one-way ANOVA with session as a factor. The average linear trend ($F_{(1,11)} = 14.3$, $p = 0.003$, $\eta^2_G = 0.428$) was significant, and the difference between sessions ($F_{(2,22)} = 3.04$, $p_{HF} = 0.097$, $\eta^2_G = 0.105$) was not significant. Therefore, response accuracy increased similarly across training sessions.

Results from the testing sessions are plotted in Figure 4.4. In both the pre- and post-training test sessions, identification accuracy was higher for 1) the full face stimulus relative to filtered stimuli; 2) horizontally filtered stimuli relative to

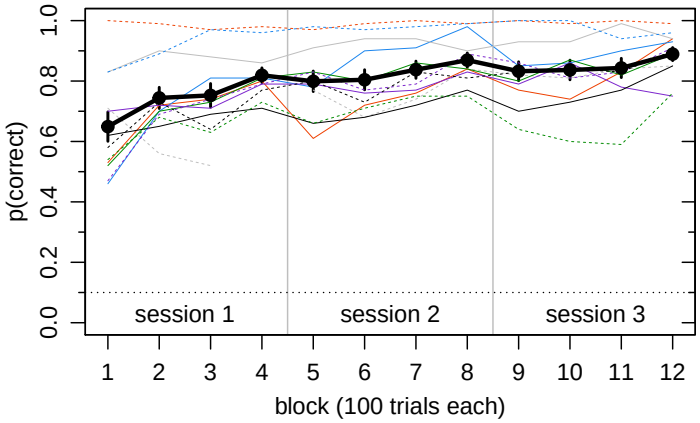

Figure 4.3 – Mean response accuracy ($+/-1$ SEM) during training, calculated for 100 trials per bin. Single subject training performance is plotted in thin coloured lines. A single subject (grey dashed) mistakenly quit the experiment after ~ 300 trials during the first session, but they returned and completed the final two sessions in total. Dashed horizontal line represents chance performance, and vertical lines separate indicate session boundaries.

vertical filtered stimuli; and 3) broader relative to narrower filter bandwidth stimuli. A comparison of Figures 4.4A and 4.4B suggests that post-training accuracy was higher than pre-training accuracy at most bandwidths. Response accuracy from testing was submitted to a 2 (*filter orientation*) × 4 (*bandwidth*) × 2 (*training*) within-subjects ANOVA. Only filter bandwidths up to ±40 deg were included, as ±40 deg is the largest bandwidth used in which the horizontally and vertically oriented filters were orthogonal. All broader bandwidths contained overlapping structures between horizontal and vertical filter orientations. The main effect of *filter orientation* ($F_{(1,11)} = 11.8$, $p = 0.0055$, $\eta^2_G = 0.190$) was significant, reflecting higher accuracy for horizontally ($M_H = 0.33$) than vertically ($M_V = 0.29$) filtered stimuli. The main effect of *bandwidth* ($F_{(3,33)} = 42.3$, $p_{HF} < 0.001$, $\eta^2_G = 0.378$) was significant, reflecting a monotonic increase in accuracy as bandwidth increased from ±10 deg to ±40 deg (accuracy increased from 16% to 38%). The main effect of *training* ($F_{(1,11)} = 90.6$, $p < 0.001$, $\eta^2_G = 0.219$) also was significant, reflecting overall higher accuracy in the post-training ($M_{post} = 0.33$) than the pre-training ($M_{pre} = 0.22$) session. Additionally, the *training* × *bandwidth* interaction ($F_{(3,33)} = 3.83$, $p_{HF} = 0.019$, $\eta^2_G = 0.024$) was significant, reflecting a larger improvement from pre- to post-training for larger bandwidths: as bandwidth increased, changes in response accuracy increased from 0.06 to 0.12. The *filter orientation* × *bandwidth* interaction ($F_{(3,33)} = 3.83$, $p_{HF} = 0.019$,

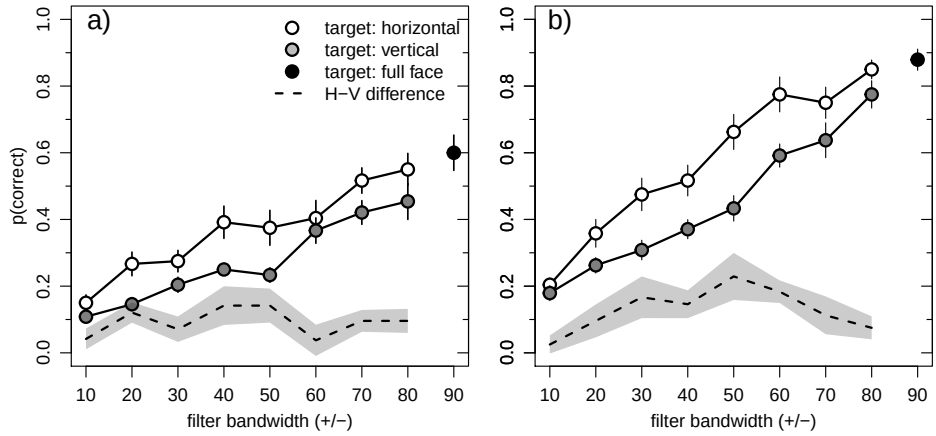

Figure 4.4 – Response accuracy in the pre-training **(A)** and post-training **(b)** test sessions. In each plot, mean response accuracy ($+/-1$ SEM) is shown for each filter bandwidth in the horizontal filter (white), vertical filter (grey), and full-face (black) conditions. Accuracy for vertical filters was subtracted from the appropriate horizontal filter conditions and plotted as the dashed line (shaded region is $+/-1$ standard error of the difference). The critical bandwidth is the $+/-40$ deg filter, as it is the largest bandwidth in which the horizontal and vertical filters contain completely orthogonal orientations without any overlap. The horizontal-vertical differences were used to calculated horizontal bias scores per participant.

$\eta^2_G = 0.037$) was significant, reflecting larger differences between horizontally and vertically filtered stimuli for certain bandwidths. The *filter orientation* × *training* interaction ($F_{(1,11)} = 0.21$, $p = 0.66$, $\eta^2_G = 0.001$), and the three-way *filter orientation* × *bandwidth* × *training* interaction ($F_{(3,33)} = 1.48$, $p_{HF} = 0.238$, $\eta^2_G = 0.013$) were not significant.

The dashed line in Figure 4.4 reflects the difference in response accuracy between horizontal and vertical conditions, calculated at each bandwidth. As found by Pachai et al. (2017), the difference in scores were approximately a quadratic function of bandwidth. This type of trend is expected if subjects have a horizontal bias, since the mid bandwidth (±40 deg) is the largest bandwidth in which the horizontal and vertical conditions differ completely. At all larger bandwidths, the two types of stimuli become increasingly similar, so identification accuracy also becomes increasingly similar. Furthermore, Pachai et al. (2018) showed that observers need a broad bandwidth (±35 deg) of horizontal structure for accurate face identification, despite the fact that most diagnostic structure is in a narrow 10 deg bandwidth. We analyzed the horizontal-vertical differences by computing the quadratic trend score for each participant, separately for pre- and post-training. Increasing horizontal bias should

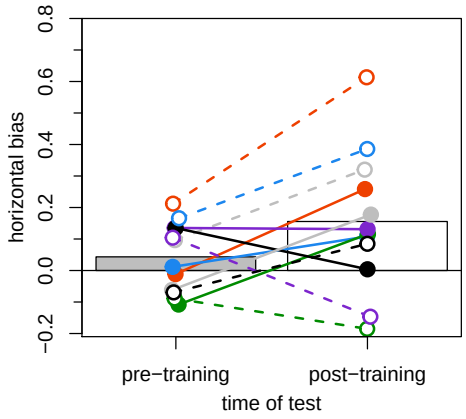

Figure 4.5 – Horizontal bias was measured at pre- and post-training sessions. Bars represent group means, and individual points/lines correspond to individual subjects. Subject colours are consistent with Figure 8. Horizontal bias was measured by first subtracting response accuracy to vertical conditions from horizontal conditions, and then computing a quadratic trend of the difference across all 8 filter bandwidths (ignoring full face). A quadratic trend (horizontal bias) score above zero represents a horizontal advantage in response accuracy that peaked at mid-size bandwidths.

increase the strength of the quadratic trend. Hence, if training increases horizontal bias in older adults, then the quadratic trend scores should be more positive after training than before training.

In a sense, the described quadratic trend score is a better measure of horizontal bias than the horizontal-vertical difference at the largest orthogonal bandwidth (e.g., ±40 deg; Chapters 2 and 3) because it is calculated using data from all bandwidths rather than a single bandwidth. If quadratic trend scores are measuring the same phenomenon as horizontal bias scores calculated using previous techniques, the two measures should be significantly correlated. Indeed, the correlation between quadratic trend scores and horizontal bias was significant at pre-training ($r = 0.795$, $t_{10} = 4.15$, $p = 0.002$) and post-training ($r = 0.903$, $t_{10} = 6.63$, $p < 0.0001$), suggesting that if quadratic trend scores are not equivalent to horizontal bias scores, they are at least tightly linked to each other.

Quadratic trend scores for pre- and post-training are plotted in Figure 4.5. On average, pre-training quadratic trend scores were near zero, and post-training quadratic trend scores were greater than zero. One-tailed t tests on quadratic trend scores confirmed that pre-training scores ($M = 0.043$, $t_{11} = 1.36$, $p = 0.101$, $d = 0.541$) were not significantly greater than zero, but post-training scores ($M = 0.156$, $t_{11} = 2.43$,

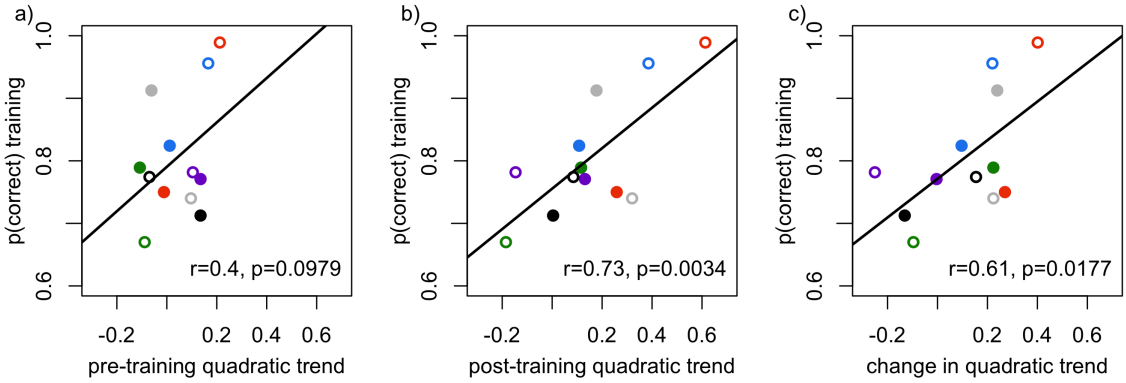

Figure 4.6 – Full face identification accuracy (y-axes) plotted against quadratic trend scores (x-axes). Full face identification accuracy calculated across all training trails plotted against quadratic trend scores measured at **A)** pre-training, **B)** post-training, and **C)** the pre- to post-training difference. Pearson's least squares correlations, and the corresponding correlation coefficients and one-tailed p-values are provided.

$p = 0.0167$, $d = 0.970$) were significantly greater than zero, and also greater than pre-training quadratic scores ($M_D = 0.112$, $t_{11} = 2.01$, $p = 0.0348$, $d = 0.580$). These results indicate that training older adults to identify full faces improved their response accuracy most to the mid-size bandwidths of horizontal filter stimuli.

Face identification accuracy was predicted to be positively correlated with quadratic trend scores, so one-tailed p-values are reported. Face identification accuracy across all training trials was not significantly correlated with pre-training quadratic trend scores ($r = 0.401$, $t_{10} = 1.39$, one-tailed $p = 0.098$, Figure 4.6A), but it was significantly correlated with post-training quadratic trend scores ($r = 0.732$, $t_{10} = 3.40$, one-tailed $p = 0.0034$, Figure 4.6B) as well as the the pre- to post-training change in quadratic trend scores ($r = 0.609$, $t_{10} = 2.43$, one-tailed $p = 0.018$, Figure 4.6C). We also inspected the correlations by correlating pre- and post-training quadratic trends with face identification accuracy separately for each block of training. When face identification accuracy was calculated separately for 12 training bins (Figure 4.7), pre-training quadratic trend scores were significantly correlated with accuracy only in the first bin ($r = 0.54$, $p = 0.034$ for block 1, and $r \leq 0.47$, $p \geq 0.07$ for blocks 2-12), whereas, post-training quadratic trend scores were significantly correlated with accuracy in all twelve bins ($r \geq 0.53$, $p \leq 0.04$ for blocks 1-12).

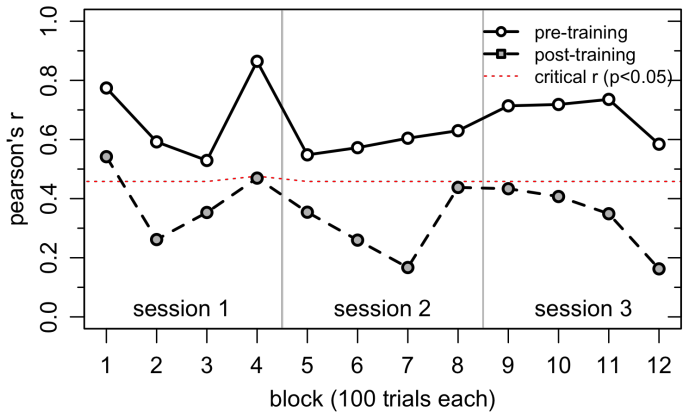

Figure 4.7 – Correlation coefficients for the association between face identification accuracy and pre-training (grey, dashed) and post-training (white, solid) quadratic trend scores. Face identification accuracy was calculated separately for each block of 100 trials. Red dotted line represents the critical r value (one-tailed $p < 0.05$; df=10 for all blocks except block 4 (df=9). Post-training quadratic trend scores were significantly correlated with face identification accuracy throughout all training blocks, whereas pre-training quadratic trend scores were significantly correlated with only the first training block.

4.5 Discussion

We trained older adults in a 1-of-10 face identification task. Response accuracy improved significantly throughout training. Before and after training, we measured horizontal bias of older adults. Horizontal bias at pre-training was negligible and not correlated with overall face identification accuracy. Critically, post-training horizontal bias was significant, and positively correlated with full face identification accuracy. Our results suggest that older adults deficits in face identification may be linked to a weakened horizontal bias, but that improved horizontal bias underlies improvements seen in training.

Post-training horizontal bias in older adults was very similar to horizontal bias to unfamiliar faces in younger adults (Pachai et al., 2017, 2018). That is, the bandwidth at which the horizontal-vertical difference peaked was ±50 deg, similar to the previously reported peaks at ±45 deg (Pachai et al., 2017, 2018). The small difference in peak bandwidth between these studies is likely due to the fact that we did not have a ±45 deg filter bandwidth in our study. Our results are consistent with the idea that human observers pool information over a broad bandwidth. For example, using the same stimulus type as here, Pachai et al. (2018) revealed that human observer performance

in a 10-AFC task is best modelled by pooling information within ±35 deg of horizontal, despite the fact that most diagnostic information is conveyed by a ±5 deg horizontal orientation band. This suggests that face identification decisions are informed by broadband orientation channels. Horizontal-vertical differences peak at narrower bandwidths for familiar than unfamiliar faces (Pachai et al., 2017), consistent with the idea that the bandwidth of orientation channels can adjust/narrow to match the stimulus (Taylor et al., 2014) with training. Results from other studies indicate that human observers pool horizontal facial structure within a much narrower orientation bandwidth (±12.5 deg: Goffaux and Greenwood, 2016), which may also be due to a narrowing of the adjustable bandwidth orientation channels to match the stimuli. However, Goffaux and Greenwood (2016) simply removed filtered orientations instead of replacing with a face context like that done here and in Pachai et al. (2018). Hence, it is possible that observers internal orientation filters were better able to adjust/narrow to stimuli containing structure in limited orientation bands, or that observers were also pooling across a broad ±35 bandwidth, but that broadband pooling does not hurt performance when orientations outside of the signal's bandwidth are left empty. One way to investigate the difference between Pachai et al. (2018) and Goffaux and Greenwood (2016) is to still embed informative horizontal structure within an uninformative face context, but leave an orientation band between the horizontal and vertical bands empty. The influence of oblique orientations on horizontal bias can be tested by varying the size of the empty bandwidth.

Full face identification accuracy was not correlated with pre-training horizontal bias, but it was significantly correlated with post-training horizontal bias. This result is consistent with the notion that perceptual learning of faces enhances older adults' sensitivity to horizontal facial structure. Similar relationships between horizontal bias and upright face identification accuracy of untrained faces have been reported in younger adults (Pachai et al., 2013; Chapter 2). Although we show a significant increase in horizontal bias from pre- to post-training, we do not have evidence that the same older adults will show an increased horizontal bias to an untrained set of faces. In fact, horizontal bias within the same set of younger adults manifests via significantly stronger horizontal bias to familiar relative to unfamiliar faces (Pachai et al., 2017), and perceptual learning of faces is stimulus-specific (Hussain et al., 2012b), indicating that the strength of horizontal bias is at least partly stimulus-specific. Generalization of a strengthened horizontal bias is critical to creating a viable face-training regimen to aide

those with deficits in horizontal bias. One possibility is to introduce adequate stimulus variability across trials in order to prevent perceptual learning's stimulus-specificity (Hussain et al., 2012a).

One explanation for the mechanism underlying the training-induced increase in horizontal bias is that that face identification judgments are made using a decision variable that relies on weighted input from lower level orientation channels, and that improvements in response accuracy are a result of the decision variable relying more on horizontally than vertically oriented channels (Figure 4.1). This interpretation is consistent with a re-weighting model of perceptual learning (Petrov et al., 2005) in which re-weighting increases the signal-to-noise ratio through increasing the weights of the signal (Gold et al., 1999b). The results can also be interpreted within the reverse hierarchy model of perception (Ahissar and Hochstein, 2004): pre-training face identification relies on a higher-level signal in which information is pooled across many orientations, and practice identifying faces manifests as a post-training face identification strategy which relies on a lower-level signal in which information is pooled across fewer orientations, and therefore informative horizontal and non-informative vertical structures can be separated. Other results suggest that the decision variable that relies most on horizontal structure may exist in the fusiform face area (Goffaux et al., 2016), further strengthening the role of a horizontally biased mechanism underlying face perception.

We show that older adults' deficits in horizontal bias (Obermeyer et al., 2012; Goffaux et al., 2015) corresponds to their overall worsened face identification accuracy. Practice in identifying a set of faces improves both face identification accuracy and horizontal bias, suggesting that deficits in face perception are likely due to a weakened horizontal bias to all faces. However, given that horizontal bias in younger adults is stronger for personally familiar than unfamiliar faces (Pachai et al., 2017), it is possible that older adults retain a strong horizontal bias for familiar faces, and they are merely less effective at identifying novel faces. This is especially likely since face identification accuracy rapidly increased after only 100 training trials (Figure 4.3); a rapid 'recovery' of horizontal bias may be indicative of an intact horizontal bias process that is only active for familiar faces. To better address this, we need to evaluate older adults' horizontal bias to personally familiar faces, as well as track the time-course of horizontal bias throughout training with unfamiliar faces. Future studies should build on the finding that horizontal bias to faces can be strengthened, and focus on

the generalizability of the improvements to novel faces.

References

Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10):457–464.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3):379–384.

Balas, B. J., Schmidt, J., and Saville, A. (2015). A face detection bias for horizontal orientations develops in middle childhood. *Frontiers in Psychology*, 06(June):1–8.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–6.

Dakin, S. C. and Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198.

Gaspar, C. M., Sekuler, A. B., and Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research*, 48(28):2817–26.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., and Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, 81:1–11.

Goffaux, V. and Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, 6:34204.

Goffaux, V., Poncin, A., and Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 Year Old). *PLoS ONE*, 10(9):1–17.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39:3537–3560.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999b). Signal but not noise changes with perceptual learning. *Nature*, 402:176–8.

Grady, C. L., Ma Maisog, J., Horwitz, B., Ungerleider, L. G., Mentis, M. J., Salerno, J. A., Pietrini, P., Wagner, E., Haxby, J. V., Gillette, J., Giacometti, K., Baldwin, P., Jacobs, G., Stein, S., Green, S., Fluck, S., and Der, M. (1994). Age-related Changes in Cortical Blood Flow Activation during Visual Processing of Faces and Location. *The Journal of Neuroscience*, 14(3):1450–1462.

Hussain, Z., Bennett, P. J., and Sekuler, A. B. (2012a). Versatile perceptual learning of textures after variable exposures. *Vision Research*, 61:89–94.

Hussain, Z., McGraw, P. V., Sekuler, A. B., and Bennett, P. J. (2012b). The rapid emergence of stimulus specific perceptual learning. *Frontiers in Psychology*, 3(July):226.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009a). Contrast-reversal abolishes perceptual learning. *Journal of Vision*, 9:1–8.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009b). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, 49(18):2273–84.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2011). Superior Identification of Familiar Visual Patterns a Year After Learning. *Psychological science*, 22(May):724–30.

Kaufmann, J. M., Schweinberger, S. R., and Burton, A. M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of cognitive neuroscience*, 21(4):625–41.

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., Broussard, C., and Others (2007). What's new in Psychtoolbox-3. *Perception*, 36(14):1.

Konar, Y., Bennett, P. J., and Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision Research*, 88:38–46.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(NOV):1–12.

Mathworks (2007). MATLAB version 7.5 (R2007b).

Nasreddine, Z., Phillips, N., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J., and Chertkow, H. (2005). The Montreal Cognitive Assessment , MoCA : A Brief Screening. *Journal of the American Geriatric Society*, 53:695–699.

Obermeyer, S., Kolling, T., Schaich, A., and Knopf, M. (2012). Differences between Old and Young Adults' Ability to Recognize Human Faces Underlie Processing of Horizontal Information. *Frontiers in Aging Neuroscience*, 4(April):3.

Olejnik, S. and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, 8(4):434–447.

Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2018). The bandwidth of diagnostic horizontal structure for face identification. *Perception*, 47(4):397–413.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to information conveyed by horizontal contours is correlated with face identification accuracy. *Frontiers in Psychology*, 4(74).

Pachai, M. V., Sekuler, A. B., Bennett, P. J., Schyns, P. G., and Ramon, M. (2017). Personal familiarity enhances sensitivity to horizontal structure during processing of face identity. *Journal of Vision*, 17(6):5.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics Transforming numbers into movies. *Spatial Vision*.

Petrov, A. A., Dosher, B. A., and Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, 112(4):715–743.

R Core Team and R Development Core Team, R. (2017). R: A Language and Environment for Statistical Computing.

Riesenhuber, M., Jarudi, I., Gilad, S., and Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(Suppl 6):S448–S450.

Rousselet, G. A., Pernet, C. R., Bennett, P. J., and Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, 9:98.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–6.

Tanaka, J. W., Curran, T., Porterfield, A. L., and Collins, D. (2006). Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. *Journal of cognitive neuroscience*, 18(9):1488–97.

Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2014). Evidence for adjustable bandwidth orientation channels. *Frontiers in Psychology*, 5(JUN):1–10.

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

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., Bub, D. N., and Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance*, 36(1):122–135.

Williams, C. and Henderson, J. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition*, 35(8):1977–1985.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1):141–145.

Chapter 5

Learning to discover orientation-specific structure in textures

5.1 Abstract

Perceptual learning refers to long-lasting improvements in performance that occur as a result of practice in a perceptual task. One prominent characteristic of perceptual learning is its high degree of stimulus specificity. For example, perceptual learning in a texture identification task is specific to the trained textures. In a typical texture identification task, perceptual learning occurs because observers become more sensitive to diagnostic stimulus components, but the particular components that are learned vary across observers. Here, we encouraged observers to adopt a specific processing strategy by manipulating the diagnostic orientation structure of textures, such that one broad orientation band contained informative target structure, and the orthogonal orientation band contained non-informative context structure. Participants practiced identifying textures containing the informative structure alone ("target-alone") or embedded within the non-informative context ("target+context"). We hypothesized that practice with target+context textures will encourage observers to *discover* the informative structure, and learning to discover informative structure will better generalize to different textures. The results reveal that practice with target-alone textures did not transfer to target+context textures, which is unsurprising since target-alone textures did

not promote the discovery strategy. However, practice with target+context textures generalized to familiar targets presented alone, indicating that participants could recognize the target structure from training, independent of the trained context. In both training groups, learning did not generalize to novel targets, suggesting that participants are learning orientation structures particular to the trained targets. In several experiments, we assess how generalization is affected by task difficulty, context variability, and ease-of-discovery. The results are interpreted using a re-weighting framework, and further discussed in different theories of perceptual learning, with special affinity to the reverse hierarchy model of perception.

5.2 Introduction

Perceptual learning refers to long-lasting improvements in performance that occur as a result of practice on many perceptual tasks (Sagi, 2011). Perceptual learning often is remarkably specific to the stimuli seen during training. For example, perceptual learning in vernier discrimination tasks (McKee and Westheimer, 1978) is specific to the stimulus orientation and retinal position seen during training (Poggio et al., 1992; Fahle et al., 1995); learning in grating discrimination tasks is specific to the spatial frequency and orientation of the trained stimulus (Fiorentini and Berardi, 1981); and direction discrimination is specific to the trained direction of motion (Ball and Sekuler, 1987). Stimulus-specific learning also occurs in arguably more complex situations such as texture identification tasks: for example, improvements in texture identification, which can last for at least one year (Hussain et al., 2011), are disrupted by changes in stimulus contrast polarity (Hussain et al., 2009a), orientation (Hussain et al., 2009c), and position (Bennett et al., 2015). Perceptual learning has been implicated in face perception as well. Life-long practice seeing faces has lead to expertise in face processing (Diamond and Carey, 1986; Gauthier et al., 2000), and since the majority of perceived faces are upright, the expertise is orientation-specific to upright, but not inverted, faces (Diamond and Carey, 1986; Sekuler et al., 2004). This phenomena is known as the face inversion effect (Yin, 1969). Inversion effects have also been observed in short-term training studies to non-face objects (Gauthier and Tarr, 1997; Rossion et al., 2002; Hussain et al., 2009a), and perceptual learning of inverted faces has even abolished the face inversion effect (Hussain et al., 2009c).

Perceptual learning occurs because practice enables observers to base their decisions

on more informative aspects of the stimuli. For example, Gold et al. (2004) showed that observers in a texture discrimination task learned to base discrimination on analyses of informative spatial regions. In other words, observers learned to discriminate textures by extracting diagnostic information from small portions of the stimuli. Similarly, the discrimination of static faces, which is affected by perceptual learning (Hussain et al., 2009b,c, 2012b; Gauthier et al., 2003), is based on diagnostic information conveyed by visual structure near the eyes and brows (Gold et al., 2004; Sekuler et al., 2004; Mangini and Biederman, 2004; Peterson and Eckstein, 2012; Rousselet et al., 2014). Discrimination of inverted faces is based on diagnostic information conveyed by the same visual structures as in upright faces, but the information is processed less efficiently (Sekuler et al., 2004; Pachai et al., 2013). Perceptual learning of inverted faces increases the efficiency of processing the most diagnostic structure (Pachai et al., 2018a), consistent with the notion that perceptual learning leads to more efficient processing (Gold et al., 1999).

Why is perceptual learning often stimulus-specific? One suggestion by Sagi (2011) is that observers learn idiosyncratic features of a response to the (often very small) set of stimuli, and their learned representations are faithful to the trained stimuli. Hence, much like a statistical algorithm can overfit data and therefore fail to account for slightly different data, observers may not be able to map similarities in stimuli that are slightly different than their learned representations. Hussain et al. (2012a) examined this idea by measuring perceptual learning in a texture identification task in conditions that differed in the amount of variation among stimuli presented during training. For example, one condition use a fixed set of stimuli whereas another condition used novel stimuli on every trial, which made it impossible to do the task based on information conveyed by a small set of exemplars. The result was that learning in the variable condition was much slower than in the fixed condition, but variable learning generalized to new items whereas learning in the fixed condition did not.

In the Hussain et al. (2012a) study, observers could only learn to perform the identification task in the variable condition by developing a strategy that could be applied to the general class of band-limited textures that were used as stimuli. Hence, it is not surprising that they could generalize learning to new textures. In the current experiments we take a different approach by examining generalization of learning in an identification task which uses a small number of fixed items during training, but the diagnostic information is embedded in a non-informative context. Our working

hypothesis is that presenting observers with a situation in which they must *discover* the task-relevant information will lead to greater generalization of learning. Specifically, this approach may reduce the likelihood that observers will learn idiosyncrasies tied to a small set of stimuli, while still using only a small set of stimuli in training. This approach is reminiscent of some practical situations where expert observers must detect a subtle signal, such as early signs of a tumour, in an unstructured and non-informative context (see Norman et al. (1989) and Nodine et al. (1999) for examples in dermatology and mammography, respectively). Furthermore, discovering the most informative structure may be the learned strategy underlying face expertise, wherein observers have learned to discover the horizontal structure of a face (Dakin and Watt, 2009; Pachai et al., 2013), which conveys important identity-related features such as the eyes and brows.

The approach to designing stimuli that require discovery of the diagnostic structure was taken from recent research on face perception. Face identification relies on a horizontal orientation band (Dakin and Watt, 2009) which conveys the most informative information about human face identity (Pachai et al., 2013). One way that horizontal selectivity has been assessed is by combining informative structure from the horizontal band of a target face with the structure from the vertical band of a non-informative face context (Pachai et al., 2018b). The result is a stimulus that appears like a full face with no visual cue about the manipulation of which orientations convey informative or non-informative structure. Therefore, identification accuracy depends on the observer's ability to accurately rely on the more informative orientation band. In the experiments presented here, stimuli were constructed using textures (e.g., Gold et al., 1999, and Figure 5.1A) instead of faces. Textures were filtered to contain informative structure in a certain orientation band (i.e., target), and combined with non-informative structure in the orthogonal orientation band (i.e., context). These stimuli, referred to as target+context stimuli, were used in a texture identification task, and were predicted to promote the discovery strategy mentioned earlier. In Experiments 1 and 2, the context is constant across items and trials (Figure 5.1 D & E), whereas in Experiments 3-6, the context varies across items and trials (e.g., Figure 5.1 F & G). In contrast, textures were also used that contained informative structure in an orientation band, and no structure in the orthogonal orientation band. These stimuli, referred to as target-alone stimuli (Figure 5.1 B & C), were predicted to not require any discovery of informative structure. If practice with target+context stimuli

promotes a discovery of the informative structure, then we should expect post-training generalization to the same targets when presented alone. Since there is no opportunity for discovery with target-alone stimuli, we do not expect practice with target-alone stimuli to generalize to the same targets when presented within a context.

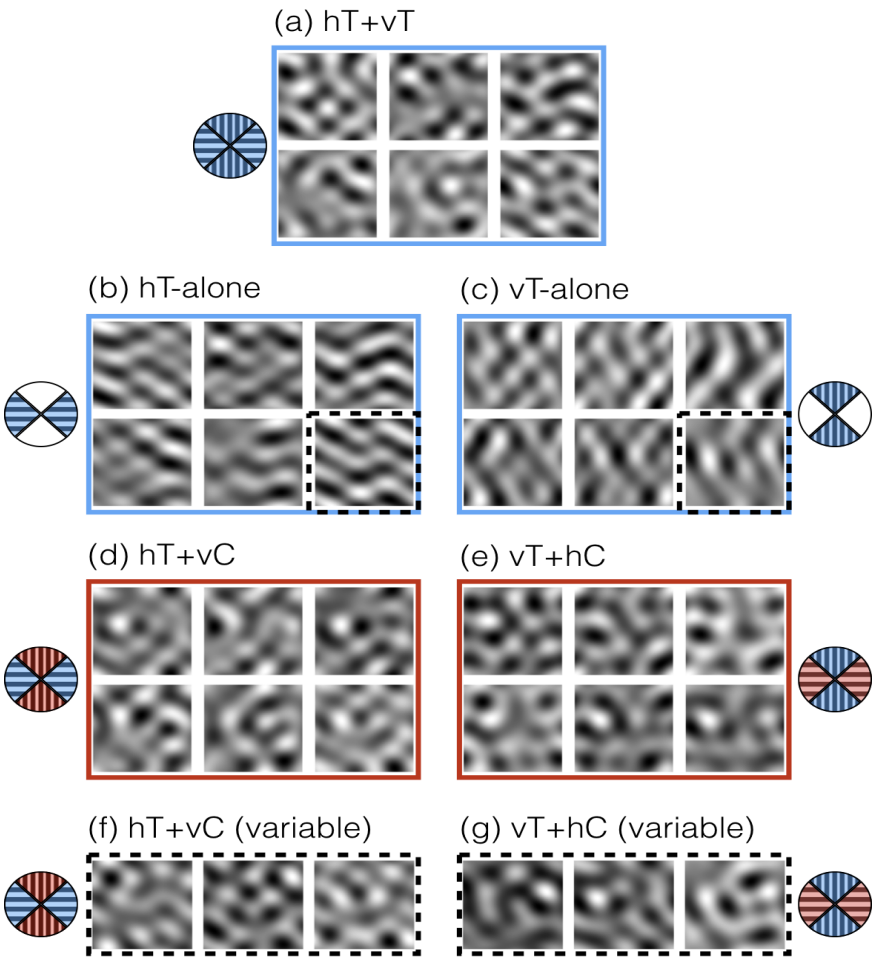

Figure 5.1 – Example sets of stimuli used: **A)** $hT + vT$: textures formed from 2-4 cycles per image spatial frequency band-limited white noise patches. They contain unique structure in all orientations (i.e., [h]orizontal and [v]ertical orientations contained diagnostic [T]arget structures, as indicated by the blue stripes in the icon). **B)** hT-alone: textures containing only horizontal structure of the textures from A, and no vertical structure. **C)** vT-alone: same as B but horizontal and vertical swapped. **D)** hT+vC: the hT-alone targets from B are each combined with the same [v]ertical [C]ontext; therefore, all six targets are identical in their vertical contexts but differ in their horizontal targets. Manipulation is indicated in the icon with the blue target and red context areas. **E)** $vT + hC$: same as D but with horizontal and vertical swapped. **F)** $hT + vC$ (variable): Example stimuli where the all three textures contain the same horizontal target (the target identified with the dotted black line from B), but a different vertical context in each exemplar. The final percept of these textures with a variable context demonstrates how different the textures appear, even though the horizontal structure is the same in all three. **G)** $vT + hC$ (variable): same as F but with a horizontal and vertical swapped. In Experiments 1 and 2, target+context textures refer to D and E, whereas in Experiments 3-6, the target+context textures always contained novel contexts on all presentations, as demonstrated in F and G.

5.3 Experiment 1: The effect of context on texture identification

5.3.1 Methods

5.3.1.1 Participants

Twenty-four undergraduate students (7 males) from McMaster University participated in the experiment. Participants had normal or corrected-to-normal visual acuity; their ages ranged from 18 to 27 years ($M = 20.5$, $SD = 3.15$). Two participants from the target+context group failed to exhibit any learning during the training phase – i.e., the difference between accuracy in the last and first two bins was ≤ 0 – and therefore were excluded from the analysis. This criterion for excluding participants, which we used in subsequent experiments, seemed reasonable because the primary aim of the experiment was to investigate the factors that influence the generalization of learning. The experimental protocol was approved by the McMaster University Research Ethics board, and informed consent was obtained from each participant prior to the experiment. Each participant received partial course credit or was paid $10/hour for their participation.

5.3.1.2 Apparatus & Stimuli

Stimuli were generated on an Apple Macintosh G5 using Matlab R2014b and the Psychophysics and Video toolboxes (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 19 inch NEC MultiSync FE992 CRT monitor, with a resolution of 1024×768 pixels (28 pixels/cm), a frame rate of 85 Hz, and an average luminance of 74.7 cd/m^2. Viewing was binocular through natural pupils, and a chin rest stabilized the viewing distance at 60 cm. The monitor was the only source of illumination in the room.

Stimuli were band-limited noise textures that were created by first applying an isotropic, band-pass (2-4 cycles/image) ideal spatial frequency filter to six 256 × 256 pixel (4.6 × 4.6 degrees of visual angle) images of Gaussian white noise. These textures were used to construct the targets for our identification tasks: they contained unique (i.e., target-specific) structure at all orientations, and therefore are referred to as horizontal-target + vertical-target ($hT + vT$) stimuli (Figure 5.1A). These textures were then passed through an ideal orientation filter that had a full bandwidth of 90°

and was centered either at 0° (i.e., horizontal) or 90° (i.e., vertical). The resulting images contained only horizontal or vertical structure and therefore we refer to them as horizontal target-alone (hT-alone) and vertical target-alone (vT-alone) stimuli (Figure 5.1B & C, respectively). Additionally, we created textures that contained diagnostic target structure in one orientation band and non-diagnostic contextual structure in the orthogonal orientation band. Vertical contextual structure (vC) was created by averaging six vT-alone textures: adding this single image to the six (hT-alone images resulted in six new textures ($hT + vC$) that contained structure at all orientations, but diagnostic (i.e., target-specific) structure only in the horizontal band (Figure 5.1D). Similarly, horizontal contextual structure (hC), created by averaging six (hT-alone textures, was added to vT-alone images to create six $vT + hC$ stimuli that had diagnostic structure only in the vertical orientation band (Figure 5.1E). During training, stimulus RMS contrast was 0.035, which is well above detection threshold. To avoid possible ceiling effects on response accuracy, stimulus RMS contrast was reduced to 0.025 during the test phase.

5.3.1.3 Design & Procedure

At the start of the experiment, participants viewed a blank grey screen for 60 s to adapt to the average luminance of the display. Each trial consisted of a 0.5 s presentation of a high-contrast fixation in the center of the display followed by a 0.15 s presentation of a single randomly-selected target texture. A response screen consisting of six textures was presented 0.5 s after target offset, and participants were instructed to use the mouse to select the item on the response screen that matched the target texture. After each response, the response screen was replaced by a blank display and participants received auditory feedback in the form of brief high- or low-pitched tones for correct and incorrect responses, respectively. The next trial started 0.2 s after the auditory feedback. Observers were instructed to select the alternative that looked like the target. The task instructions were identical for the testing and training sessions.

Half of the participants were trained on six hT-alone (i.e., target-alone) textures, while the other half trained on six $hT + vC$ (i.e., target+context) textures. Training consisted of 960 trials total (160 repetitions per target), split equally across two consecutive days. There also were two testing sessions split across two days: On Day 1, the testing session occurred prior to the training trials, and on Day 2 the testing session occurred after the training trials. Each testing session had two blocks. Block A

consisted of hT_F-alone, vT_N-alone, and $hT_F + vT_N$ stimuli, where the subscripts F and N denote Familiar and Novel target textures, respectively. Testing block B consisted of $hT_F + vC$, $vT_N + hC$, and $hT_F + vT_N$ stimuli. In each testing block, each stimulus was presented 60 times in a random order, yielding a total of 180 trials per block. The order of blocks A and B were randomized across participants, but within each participant the same order was used on days 1 and 2. The conditions are summarized in Table 5.1.

As noted above, during each testing phase subjects were presented with stimuli that were designated as familiar or novel. A stimulus was designated as familiar if it corresponded to a pattern that was seen during the training phase (Table 5.1). For example, hT_F-alone would correspond to a horizontally-filtered texture that was presented during training, whereas $hT_F + vT_N$ would be a texture that consisted of a familiar, horizontally-filtered texture *and* a novel, vertically-filtered texture (Figure 5.1B&A). The vT_N-alone and $vT_N + hC$ stimuli were novel in two respects. First, these filtered stimuli were constructed from white noise textures that were different from the ones used to generate the horizontal patterns seen during training. Second, the informative parts of the textures were oriented vertically instead of horizontally (Figure 5.1C&E). Based on the findings of Hussain et al. (2009a), we expected that the effects of practice with horizontal textures would not transfer to novel, vertical textures. We can therefore use improvements with the vertical target textures (vT_N-alone and $vT_N + hC$) as a measure of learning that is unrelated to the stimulus, and include these conditions in our ANOVAs and follow-up t-tests as a standard for measuring improvements in the horizontal target conditions.

Day 1			
Testing Block A	hT_F-alone	vT_N-alone	$hT_F + vT_N$
Testing Block B	$hT_F + vC$	$vT_N + hC$	$hT_F + vT_N$
Training	6 hT-alone textures	or 6 $hT + vC$ textures	
Day 2			
Training	6 hT-alone textures	or 6 $hT + vC$ textures	
Testing Block A	hT_F-alone	vT_N-alone	$hT_F + vT_N$
Testing Block B	$hT_F + vC$	$vT_N + hC$	$hT_F + vT_N$

Table 5.1 – Experimental conditions in Experiment 1. The order of Blocks A and B was randomized for each participant but was the same on both days. During training, half of the subjects saw hT-alone textures and half saw $hT + vC$ textures.

5.3.2 Results

All statistical analyses were performed with R (R Core Team and R Development Core Team, 2017). For F tests, effect size was measured using generalized eta squared (η^2_G) because it provides comparable values for between- and within-subject measures (Bakeman, 2005; Olejnik and Algina, 2003). Lakens (2013) suggests that Cohen (1988)'s suggested benchmarks of effect size can be used when interpreting η^2_G (small: 0.01, medium: 0.04, large: 0.16). Unless stated otherwise, the reported t tests and p values are all two-tailed, and are accompanied by mean difference values (M_D). Additionally, for the reader's benefit, F tables for most of the described ANOVAs on the testing data are reproduced in full in the Appendix.

5.3.2.1 Training Phase

Response accuracy during training trials is shown in Figure 5.2A. Overall response accuracy during training was significantly better for the target-alone trained group compared to the target+context group ($M_D = 0.107$, $t_{20} = 3.68$, $p = 0.001$). Training data were split into 10 bins of 96 trials each, and a preliminary trend analysis indicated that response accuracy increased approximately linearly across the 10 training trial bins, and that the higher-order trends were small and non-significant (F's< 1). Therefore our analyses of the training data focussed on the linear trend of proportion correct across the 10 training trial bins. We found that the linear trend was significant ($F_{(1,20)} = 165.2$, $p < 0.0001$, $\eta^2_G = 0.892$) and that the linear trend did not differ across groups ($F_{(1,20)} = 1.40$, $p = 0.251$, $\eta^2_G = 0.065$). Hence, the data indicate that overall accuracy during training was higher in the target-alone than the target+context group, but that learning occurred in both groups and that the rate of learning, as indexed by the linear trend, did not differ between groups.

5.3.2.2 Pre- & Post-training Test Phases

Differences between pre- and post-training scores in each test condition are shown in Figure 5.2B. Data from the hT_F-alone, $hT_F + vC$, vT_N-alone, and $vT_N + hC$ conditions were analyzed with a 2 (*training condition*) × 2 (*testing context*) × 2 (*stimulus novelty*) ANOVA, which revealed a significant three-way interaction between *training condition* (target-alone vs. target+context), *testing context* (target-alone vs. target+context), and *stimulus novelty* (familiar horizontal vs. novel vertical targets) ($F_{(1,20)} = 44.3$,

$p < 0.0001$, $\eta^2_G = 0.212$). Therefore we analyzed the data from each training group with separate 2 (*testing context*) × 2 (*stimulus novelty*) within-subject ANOVAs.

For the target-alone trained group (white bars in Figure 5.2B), the ANOVA revealed a significant intercept ($F_{(1,11)} = 44.5$, $p < 0.0001$, $\eta^2_G = 0.534$), indicating that, overall, accuracy was higher after training. The main effect of *testing context* ($F_{(1,11)} = 8.17$, $p = 0.0156$, $\eta^2_G = 0.179$) was significant, but the main effect of *stimulus novelty* ($F_{(1,11)} = 2.14$, $p = 0.171$, $\eta^2_G = 0.050$) was not. The *testing context* × *stimulus novelty* interaction ($F_{(1,11)} = 11.6$, $p = 0.0059$, $\eta^2_G = 0.136$) was significant, and follow-up paired t-tests (two-tailed) showed that accuracy increased more for trained textures than untrained textures that were presented alone (i.e., hT_F-alone > vT_N-alone; $M_D = 0.136$, $t_{11} = 3.50$, $p = 0.006$), but not when the textures were embedded within a non-informative context (i.e., $hT_F + vC \not> vT_N + hC$; $M_D = -0.036$, $t_{11} = -0.788$, $p = 0.447$). These results suggest that training with horizontal target-alone textures resulted in stimulus specific learning (hT_F-alone > vT_N-alone) that did not generalize to a situation where the familiar horizontal targets were embedded in a novel context ($hT_F + vC \not> vT_N + hC$).

For the target+context trained group (grey bars in Figure 5.2B), the ANOVA found that the intercept was significantly different from zero ($F_{(1,9)} = 14.1$, $p = 0.0045$, $\eta^2_G = 0.341$), indicating that, like what was found for the target-alone group, overall accuracy in the test conditions was higher after training. As was found for the target-alone group, the main effect of *testing context* ($F_{(1,9)} = 10.65$, $p = 0.0010$, $\eta^2_G = 0.227$) was significant, the main effect of *stimulus novelty* ($F_{(1,9)} = 0.0063$, $p = 0.808$, $\eta^2_G = 0.002$) was not significant, and the *testing context* × *stimulus novelty* interaction ($F_{(1,9)} = 55.0$, $p < 0.0001$, $\eta^2_G = 0.313$) was significant. The interaction reflects the fact that accuracy improved more for familiar horizontal textures than novel vertical textures when they were embedded in context (i.e., $hT_F + vC > vT_N + hC$; $M_D = 0.135$, $t_9 = 2.95$, $p = 0.0161$), which was not true for textures presented without context (i.e., hT_F-alone < vT_N-alone; $M_D = -0.117$, $t_{11} = -3.42$, $p = 0.0077$). These results show that, as was found with the target-alone trained group, training with horizontal targets embedded in a non-informative vertical context produced stimulus specific learning ($hT_F + vC > vT_N + hC$) that did not generalize to the case where the same horizontal targets were presented without a non-informative context (hT_F-alone $\not> vT_N$-alone).

Our results indicate that stimulus-specific learning was significant in both groups:

response accuracy after training was most improved when subjects were presented with the same stimulus seen during training. In addition, we found that that neither group showed evidence of context-generalization: the effects of training were significantly reduced if the familiar targets were placed in a non-informative context, or if the familiar targets were presented without the non-informative context that was present during training. In addition, the magnitude of stimulus-specific learning, which was defined in the above tests as hT_F-alone $> vT_N$-alone for the target-alone group, and as $hT_F + vC > vT_N + hC$ for the target+context group, did not differ between the two training groups ($M_D = 0.001$, $t_{20} = 0.019$, $p = 0.985$). Similarly, the magnitude of context-generalization, which was defined in the above tests as $hT_F + vC > vT_N + hC$ for the target-alone group, and as hT_F-alone $> vT_N$-alone for the target+context group ($M_D = 0.0081$, $t_{20} = 1.36$, $p = 0.189$) did not differ between the two training groups. Hence, our results suggest that the two training groups exhibited similar behaviour in the training phase.

Finally, we compared the effect of training on response accuracy in the hT_F-alone and $hT_F + vC$ testing conditions with the $hT_F + vT_N$ (full target textures as in the top of Figure 5.1A; data on far right of Figure 5.2B). $hT_F + vT_N$ textures contain diagnostic target structure in all orientations, unlike the other two hT_F textures that contain informative structure only in the horizontal orientations. Therefore, $hT_F + vT_N$ stimuli provide an opportunity to assess the effect of training on stimuli that appear like $hT_F + vC$ textures, but in which the structure at the horizontal *and* vertical orientations is informative. In the target-alone group, training resulted in a significant improvement in response accuracy with $hT_F + vT_N$ textures ($M = 0.108$, $t_{11} = 3.29$, $p = 0.0072$). This change in response accuracy was significantly smaller than improvements to hT_F-alone textures ($M_D = -0.126$, $t_{11} = -2.70$, $p = 0.0207$), and was slightly, but not significantly, larger than improvements to $hT_F + vC$ textures ($M_D = 0.061$, $t_{11} = 1.94$, $p = 0.0782$). Therefore, training in the target-alone condition had similar effects on accuracy for test stimuli in which the vertical context was or was not informative. Participants who trained on target+context textures showed improvements to $hT_F + vT_N$ textures ($M = 0.117$, $t_9 = 4.16$, $p = 0.002$) that were slightly but not significantly smaller than the training effect for $hT_F + vC$ textures ($M_D = -0.068$, $t_9 = -2.21$, $p = 0.0545$), and was significantly larger than the effect for hT_F-alone textures ($M_D = 0.158$, $t_9 = 4.91$, $p = 0.0008$). Therefore, in the target+context group, removing all vertical structure (hT_F-alone) from the trained

stimulus reduced the effect of training more than replacing the non-informative context with diagnostic structure ($hT_F + vT_N$). In both training conditions, improvements were largest when the textures contained the *same* target and context seen during training.

5.3.3 Discussion

Results from Experiment 1 suggest that training to identify horizontal textures that were or were not embedded in non-informative vertical structure produced similar learning: the rate of learning during the training phase did not differ significantly between groups, and the learning in both groups exhibited similar patterns of stimulus-specificity. Specifically, training improved response accuracy most for the trained patterns presented in the trained context. Contrary to our hypothesis, we found no evidence that embedding the diagnostic structure within a non-informative context produced greater generalization of learning.

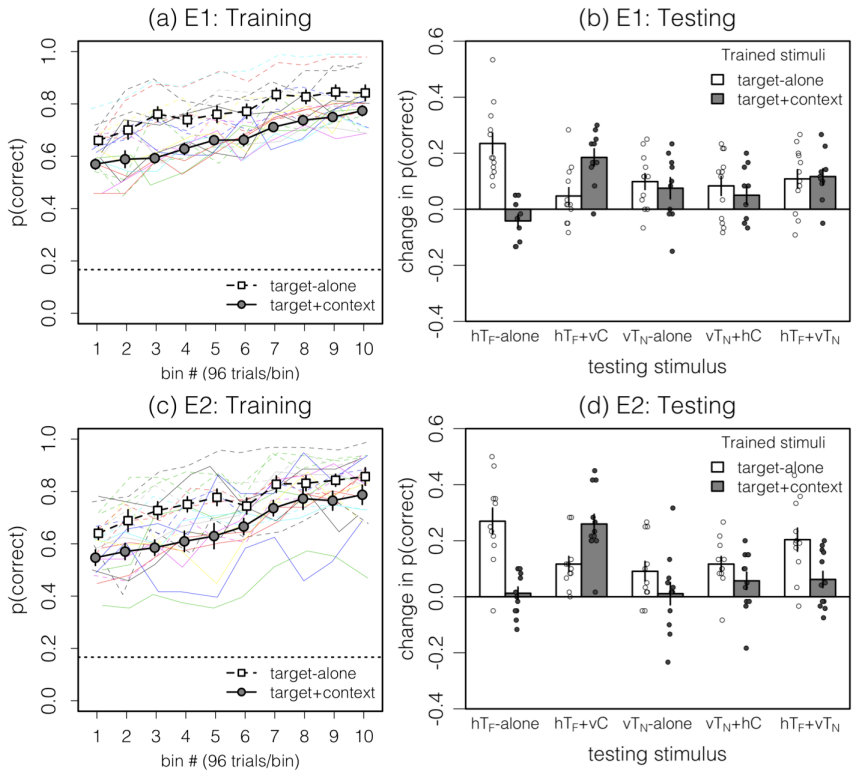

Figure 5.2 – **A)** Mean proportion correct for each bin of 96 trials across both training sessions from Experiment 1 (Day 1: bins 1-5: Day 2: bins 6-10). Dotted line represents chance performance. **B)** Change in the mean proportion correct from pre- to post-training for the different testing conditions from Experiment 1. **C & D)** Same as A & B except for Experiment 2. Note that the ANOVA performed on the test data did not include the $hT_F + vT_N$ condition. Error bars represent ± 1 SEM.

5.4 Experiment 2: The effect of target contrast on training task difficulty

One potential criticism of Experiment 1 is how we constructed our images. Our stimuli were constructed by setting the RMS contrast *after* the target and context were combined; hence, the contrast power for the diagnostic horizontal structure was lower in the target+context condition compared to the target-alone condition. In other words, stimulus contrast was set for the final texture, and therefore the target structure had a slightly higher contrast in target-alone textures than in target+context textures. The contrast difference may explain why response accuracy was higher in the target-alone group than the target+context group. Differences in difficulty are especially interesting since increasing task difficulty during training may increase the stimulus-specificity of perceptual learning (Ahissar and Hochstein, 1997). Hence, it is possible that an effect of context on learning generalization was counteracted by the effect of increased difficulty in the target+context condition. Therefore, in Experiment 2 we examined how accuracy in the training and test phases might be affected by equating RMS contrast of the diagnostic structure in the two training conditions.

5.4.1 Methods

Twenty four naïve participants (9 males, range=18-25 years old, M=20.0, SD=2.71), 12 in each training condition, were tested in this experiment. One participant from the target-alone group failed to exhibit any learning and was thus excluded from the analysis. The apparatus, procedure, and experimental design were identical to those used in Experiment 1, with the exception that the target RMS contrast was the same in the target-alone and target+context textures (0.035 during training and 0.025 during testing), and the context in the target+context stimuli had the same RMS contrast as the target.

5.4.2 Results

5.4.2.1 Training Phase

Training data is presented in Figure 5.2C. During training, mean response accuracy was significantly better in the target-alone group than the target+context group

($M_D = 0.102$, $t_{21} = 2.51$, $p = 0.020$). Proportion correct improved approximately linearly across 96-trial bins: the linear trend was statistically significant ($F_{(1,21)} = 137.2$, $p < 0.0001$, $\eta^2_G = 0.867$) and did not significantly differ between groups ($F_{(1,21)} = 2.35$, $p = 0.140$, $\eta^2_G = 0.101$).

The difference between accuracy in the two training groups was similar in Experiment 2 ($M_D = 0.102$) and Experiment 1 ($M_D = 0.107$). Mean response accuracy during the training phase was submitted to a 2 (*training group*) × 2 (*experiment*) ANOVA. The main effect of *training group* ($F_{(1,41)} = 17.2$, $p = 0.0002$, $\eta^2_G = 0.296$) was significant, but the main effect of *experiment* ($F_{(1,41)} = 0.017$, $p = 0.897$, $\eta^2_G < 0.001$) and the two-way interaction ($F_{(1,41)} = 0.009$, $p = 0.923$, $\eta^2_G < 0.001$) were not significant. Furthermore, a 2 (Experiment) × 2 (Training Group) ANOVA on the linear trend scores showed that the linear increase of accuracy across the 96-trial bins of training trials did not differ across experiments ($F_{(1,41)} = 2.28$, $p = 0.139$, $\eta^2_G = 0.053$) and training condition ($F_{(1,41)} = 3.70$, $p = 0.0615$, $\eta^2_G = 0.083$), and the difference between conditions did not vary significantly across experiments ($F_{(1,41)} = 0.248$, $p = 0.621$, $\eta^2_G = 0.006$).

Finally, we conducted a 2 (*training group*) × 2 (*experiment*) × 10 (*bin*) ANOVA and found that the the main effects of *training group* ($F_{(1,41)} = 17.2$, $p = 0.0002$, $\eta^2_G = 0.229$) and *bin* ($F_{(9,369)} = 74.2$, $p_{HF} < 0.0001$, $\eta^2_G = 0.347$) were significant, as was the *bin* × *training group* interaction ($F_{(9,369)} = 2.40$, $p_{HF} = 0.021$, $\eta^2_G = 0.017$) also was significant. All other main effects and interactions were not significant (in each case, $F< 0.8$, $p> 0.68$, & $\eta^2_G < 0.006$). Inspection of Figure 5.2 suggests that the significant *bin* × *training group* interaction reflects the fact that in both experiments the difference between training groups was slightly larger in the first half of training than the second half. We evaluated this hypothesis by computing average response accuracy in training session 1 (i.e., bins 1-5) and training session 2 (i.e., bins 6-10) for each participant, and submitting those scores to a 2 (training group) × 2 (session) ANOVA. As expected, the *training group* × *session* interaction was significant ($F_{1,43} = 6.12$, $p = 0.012$, $\eta^2_G = 0.0125$), indicating that the difference between groups was smaller in session 2 than session 1.

Overall, these analyses suggest that response accuracy during training did not differ significantly between Experiments 1 and 2: In both experiments, overall response accuracy during the training phase was significantly lower in the target+context group than the target-alone group, but that the change in accuracy, as indexed by the linear

trend, was similar in the two groups.

5.4.2.2 Pre- & Post-training Test Phases

Differences between pre- and post-training scores in each test condition are shown in Figure 5.2D. Differences between pre- and post-test scores in the hT_F-alone, $hT_F + vC$, vT_N-alone, and $vT_N + hC$ conditions were analyzed with a 2 (*training condition*) × 2 (*testing context*) × 2 (*stimulus novelty*) ANOVA, which revealed that the overall change in performance differed significantly from zero ($F_{(1,21)} = 53.9$, $p < 0.0001$, $\eta^2_G = 0.530$), indicating that performance was better, on average, after training. As was found in Experiment 1, the ANOVA also found a significant three-way interaction between *training condition*, *testing context*, and *stimulus novelty* ($F_{(1,21)} = 21.2$, $p = 0.0002$, $\eta^2_G = 0.157$), so we analyzed the data from each training group with separate 2 (*testing context*) × 2 (*stimulus novelty*) within-subject ANOVAs.

In the target-alone trained group (white bars in Figure 5.2D), the grand mean differed significantly from zero ($F_{(1,10)} = 34.4$, $p = 0.0002$, $\eta^2_G = 0.640$), indicating that, overall, accuracy in the test conditions improved with training. The main effect of *testing context* ($F_{(1,10)} = 3.92$, $p = 0.0758$, $\eta^2_G = 0.075$) was not significant, but the main effect of *stimulus novelty* ($F_{(1,10)} = 8.70$, $p = 0.0145$, $\eta^2_G = 0138$) and the *testing context* × *stimulus novelty* interaction ($F_{(1,10)} = 17.5$, $p = 0.0019$, $\eta^2_G = 0.138$) were significant. Follow-up t tests revealed that accuracy increased more for trained textures than untrained textures that were presented alone (hT_F-alone > vT_N-alone, $M_D = 0.179$, $t_{10} = 5.12$, $p = 0.0004$) but not when those textures were embedded in a non-informative context ($hT_F + vC \not> vT_N + hC$, $M_D < 0.001$, $t_{10} < 0.001$, $p > 0.99$). Hence, these results suggest that training with horizontal target-alone textures resulted in stimulus specific learning (hT_F-alone > vT_N-alone) that did not generalize to a situation where the familiar horizontal targets were embedded in a novel context ($hT_F + vC \not> vT_N + hC$).

In the target+context trained group (grey bars in Figure 5.2D), the grand mean differed significantly from zero ($F_{(1,11)} = 18.6$, $p = 0.0012$, $\eta^2_G = 0.381$), indicating that, overall, accuracy was higher after training. The main effect of *testing context* ($F_{(1,11)} = 32.7$, $p = 0.0001$, $\eta^2_G = 0.314$) and the main effect of *stimulus novelty* ($F_{(1,11)} = 11.8$, $p = 0.0055$, $\eta^2_G = 0.182$) were significant, and they were qualified by a significant *testing context* × *stimulus novelty* interaction ($F_{(1,11)} = 8.65$, $p = 0.013$, $\eta^2_G = 0.178$). Follow-up t tests indicated that the interaction reflected the fact that

accuracy improved more for the familiar horizontal textures than the novel vertical textures when the textures were embedded in non-informative context ($hT_F + vC > vT_N + hC$, $M_D = 0.203$, $t_{11} = 4.33$, $p = 0.0012$), but not when the textures were presented without a context (hT_F-alone $\not> vT_N$-alone, $M = 0.001$, $t_{11} = 0.032$, $p = 0.975$). Hence, these results show that, as was found with the target-alone trained group, training with horizontal targets embedded in a non-informative vertical context produced stimulus specific learning ($hT_F + vC > vT_N + hC$) that did not generalize to the case where the same horizontal targets were presented without a non-informative context (hT_F-alone $\not> vT_N$-alone).

As was found in Experiment 1, both training groups demonstrated significant stimulus-specific learning but no evidence for context-generalization. Furthermore, as in Experiment 1, the magnitude of stimulus-specific learning ($M_D = -0.023$, $t_{21} = -0.40$, $p = 0.69$) and context-generalization ($M_D = -0.001$, $t_{21} = -0.02$, $p = 0.98$) did not differ between training conditions.

Finally, we compared the effect of training on response accuracy in the hT_F-alone and $hT_F + vC$ testing conditions with the $hT_F + vT_N$ testing condition. In the target-alone group, the effect of training on the $hT_F + vT_N$ textures ($M = 0.204$, $t_{10} = 4.94$, $p = 0.0006$) was significant. This effect was slightly but not significantly smaller than the training effect on hT_F-alone textures ($M_D = -0.066$, $t_{10} = -2.08$, $p = 0.064$), but was larger than the training effect on $hT_F + vC$ textures ($M_D = 0.087$, $t_{10} = 3.21$, $p = 0.0092$). Therefore, unlike Experiment 1, the effect of training on familiar targets in an informative context was more similar to the the stimuli used during draining, than the stimuli with an uninformative context. In the target+context trained group, the effect of training on accuracy for the $hT_F + vT_N$ textures ($M = 0.062$, $t_{11} = 2.15$, $p = 0.055$) was not significant. This effect was smaller than the effect of training on $hT_F + vC$ textures ($M_D = -0.198$, $t_{11} = -6.87$, $p < 0.0001$), and was slightly but not significantly greater than the effect on hT_F-alone textures ($M_D = 0.049$, $t_{11} = 1.14$, $p = 0.277$). Therefore, in the target+context group, removing all vertical structure (hT_F-alone) from the trained stimulus reduced the effect of training more than replacing the non-informative context with diagnostic structure ($hT_F + vT_N$). As was found in Experiment 1, in both training conditions, improvements were largest when the textures contained the same target and context type as the textures seen during training, although, relative to Experiment 1, the target-alone group showed greater improvements to targets with an informative context.

5.4.2.3 Test Phase Data combined across Experiments 1 & 2

To examine whether the test data differed between experiments, the difference between pre- and post-training accuracy in the hT_F-alone, $hT_F + vC$, vT_N-alone, and $vT_N + hC$ conditions in both experiments were analyzed with a 2 (*experiment*) × 2 (*training condition*) × 2 (*testing context*) × 2 (*stimulus novelty*) ANOVA. The ANOVA found a significant *experiment* × *novelty* interaction ($F_{(1,41)} = 4.08$, $p = 0.05$, $\eta^2_G = 0.024$), which reflected the fact that effect of novelty was larger in Experiment 2 ($F_{(1,22)} = 21.37$, $p < .001$, $\eta^2_G = 0.21$) than in Experiment 1 ($F_{(1,21)} = 1.61$, $p = 0.21$, $\eta^2_G = 0.034$). None of the other effects involving *experiment* were significant (in each case, $F \leq 1.76$, $p \geq .19$, $\eta^2_G \leq 0.014$). As was found in the analyses of the individual experiments, the ANOVA also yielded a significant three-way interaction between *training*, *stimulus novelty*, and *testing context* ($F_{(1,41)} = 58.75$,$p < .001$, $\eta^2_G = 0.183$). This three-way interaction was analyzed by performing separate 2 (*testing context*) × 2 (*stimulus novelty*) within-subject ANOVAs on the target-alone and target+context groups while combining data across experiments: in both ANOVAs, the main effects of *testing context* and *stimulus novelty*, as well as the two-way interaction, were significant (in all cases, $F \geq 5.82$, $p \leq .025$, $\eta^2_G \geq 0.074$). The *testing context* × *stimulus novelty* interaction in both training groups, after combining data across experiments, is illustrated in Figure 5.3. Inspection of Figure 5.3 shows that the two-way interaction reflects the fact that both training groups exhibited more learning to the stimuli seen during training than to the same targets presented in a different context. In other words, the pattern of test results for the combined data was quantitatively similar to the patterns observed in the individual experiments (Cf. Figure 5.3 and Figures 5.2 B & 5.2 D).

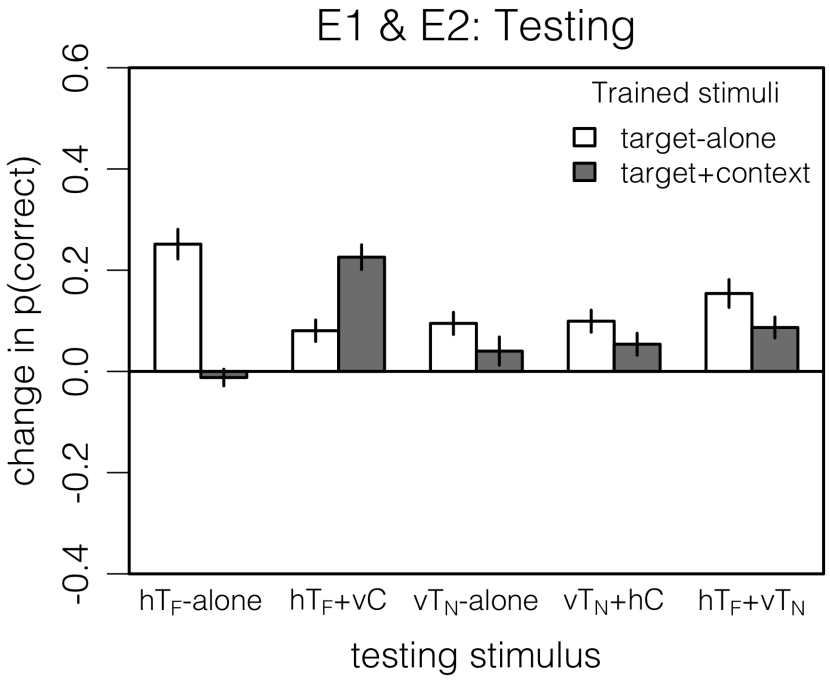

Figure 5.3 – The difference between pre- and post-training response accuracy for the test conditions in Experiments 1 & 2. The means are shown for the data combined across experiments. Error bars represent ±1 SEM.

5.4.3 Discussion

Experiment 2 examined whether the difference in task difficulty during training in Experiment 1 was due to differences in the stimulus' target contrast, and if so, if it affected the stimulus-specificity of learning in the target-alone and target+context training groups. Unlike Experiment 1, in Experiment 2 the target contrast was equated in the target-alone and target+context textures. This change in the stimulus did not significantly reduced the difference between the two training groups (cf., Figures 5.2 A & C), suggesting that differences in task difficulty were not due to target contrast differences, but were due to the status of the uninformative context. Indeed, statistical comparisons between Experiments 1 and 2 found that the main effect of experiment, as well as all of the interactions including experiment, accounted for very small amounts of variation in response accuracy (i.e., $\eta^2 < 0.006$). Unsurprisingly, the manner in which learning in the two groups generalized across several test conditions was essentially the same in Experiments 1 and 2. In particular, both experiments found that improvements in response accuracy were largest for the stimulus that used the same stimulus seen during training, and that neither training group generalized learning to the same targets presented in a new context. Specifically, participants trained with hT-alone stimuli did not generalized learning to $hT + vC$ stimuli, and participants trained with $hT + vC$ stimuli failed to generalize learning to hT-alone stimuli (see, for example, the asymmetry in the two leftmost pairs of bars in Figure 5.2 B & D). Hence, contrary to our original hypothesis, both experiments suggest that training to discriminate targets in the target+context condition does *not* lead to more generalizable learning than training in the target-alone condition.

In some respects the current results are surprising. For example, both experiments found that accuracy in the target+context condition increased with training, and that training improved observers' sensitivity to the textures's diagnostic, horizontal target structure. One simple explanation for this finding is that observers perform the texture identification task using a decision variable that is constructed by combining the responses of an ensemble of orientation-selective channels (DeValois and DeValois, 1990), and that training enables observers to increase the relative weight assigned to channels tuned to horizontal structure, which conveys diagnostic information (Dosher and Lu, 1999; Petrov et al., 2005). This simple framework predicts that learning in the target+context condition ought to generalize to the target-alone condition. However,

the results of Experiments 1 and 2 are inconsistent with this hypothesis, which fails to account for the fact that in both experiments learning in the target+context condition did not generalize to the target-alone condition, in which the same diagnostic information was presented in the absence of the non-informative context. Hence, our results suggest that the non-informative context significantly influenced performance even after training, and therefore that it is incorrect to say that participants were learning to ignore, or de-emphasize, the non-informative context during training. If the mechanism of learning is through channel re-weighting, the results suggest that the decision variable relies on the re-weighting of channels that have already combined the outputs of all, or many, orientation-selective channels (i.e., the pooled representation of the target and context, as in Olzak and Thomas, 1999).

Of course, it is possible that observers could learn to ignore the non-informative context in other situations. Suppose, for example, that observers performed the texture identification task by performing local spatial analyses on representations that combined target and context structure, and that during training they came to rely on local regions that varied significantly across textures (Gold et al., 2004). Such a strategy would be difficult to implement in situations in which the contextual structure – and therefore the overall spatial structure – varied randomly across trials. Hence, introducing trial-by-trial variation in stimulus context might force observers to use a strategy that relies solely on the target structure and, consequently, produce learning that generalizes across changes in stimulus context. The following experiments test this hypothesis.

5.5 Experiment 3: Distinguishing the target from a variable context

In Experiment 3, we altered the stimuli in ways that we hoped would discourage participants from learning texture representations that include the non-informative context. Specifically, instead of using the same, non-informative context in all textures, we appropriately filtered novel, random noise fields to create a new, non-informative contextual structure for each pattern. In other words, the texture stimulus and the six alternatives on the response screen all contained a different random context on every trial. In addition, to make it easier to distinguish the target and contextual structure,

we reduced the orientation bandwidths of the target and context from ±45 to ±30 deg and removed all frequency components with orientations of 45 ± 15 and 135 ± 15 deg.

5.5.1 Methods

The stimuli were band-pass (2-4 cyc/image) filtered textures that consisted of horizontal (0 ± 30 deg) target and vertical (90 ± 30 deg) contextual structure. Spatial frequency components at oblique orientations (45 ± 15 & 135 ± 15 deg) were removed from the images. As in previous experiments, there were six unique targets that remained constant across participants and trials. However, unlike previous experiments, new contextual structure was created for each stimulus on every trial by appropriately filtering novel, random noise fields. All other aspects of the apparatus, experimental procedure, and stimuli, including the target and context contrast, were the same as in Experiment 2.

Twenty-four naïve participants (3 males, range = 18-28 years old, M = 20.3, SD = 2.58) completed this experiment, 12 in each target-alone and target+context training group. Three participants from the target+context group failed to exhibit any learning during the training phase.

5.5.2 Results

5.5.2.1 Training Phase

The performance during the training phase is plotted in Figure 5.4A. During training, response accuracy, averaged across bins, was significantly higher in the target-alone group than in the target+context group ($M_D = 0.42$, $t_{22} = 16.7$, $p < 0.0001$). The linear trend of accuracy across bins was significant ($F_{(1,22)} = 75.8$, $p < 0.0001$, $\eta^2_G = 0.775$); however, unlike what was found in previous experiments, the linear trend significantly differed between groups ($F_{(1,22)} = 4.70$, $p = 0.041$, $\eta^2_G = 0.176$). Follow-up analyses indicated the linear trend was significant in both groups (target-alone: $M = 0.179$, $t_{11} = 7.51$, $p < 0.0001$; target+context: $M = 0.108$, $t_{11} = 4.73$, $p = 0.0003$), but was larger in the target-alone group. Hence the difference between groups in response accuracy increased across training bins.

Accuracy was compared between training conditions and between Experiments 2 and 3. The difference in accuracy between training conditions was significantly

larger in Experiment 3 than in Experiment 2 ($F_{(1,43)} = 45.9$, $p < 0.0001$, $\eta^2_G = 0.517$), reflecting the lower response accuracy for the target+context group in Experiment 3 due to the variable context. Importantly, the linear increase across the 96-trial bins of training trials differed across experiments for different training conditions ($F_{(1,43)} = 6.59$, $p = 0.0138$, $\eta^2_G = 0.13$), reflecting a smaller linear trend for the target+context group in Experiment 3 than all other groups.

5.5.2.2 Pre- & Post-training Test Phases

The difference between pre- and post-training accuracy is shown for each group and test condition in Figure 5.4B. The ANOVA on the difference scores revealed that the grand average differed from zero ($F_{(1,22)} = 60.8$, $p < 0.0001$), indicating that overall accuracy was higher after training. As was found in Experiments 1 and 2, the *training condition* × *testing context* × *stimulus novelty* interaction was significant ($F_{(1,22)} = 10.4$, $p = 0.0039$, $\eta^2_G = 0.073$), and therefore we completed the rest of our analysis using separate 2 (*testing context*) × 2 (*stimulus novelty*) ANOVAs for each training group.

In the target-alone group (white bars in Figure 5.4B), the intercept was significantly different from zero ($F_{(1,11)} = 26.9$, $p = 0.0003$, $\eta^2_G = 0.469$), indicating that accuracy was higher after training. The main effects of *testing context* ($F_{(1,11)} = 17.0$, $p = 0.0017$, $\eta^2_G = 0.326$) and *stimulus novelty* ($F_{(1,11)} = 20.0$, $p = 0.0009$, $\eta^2_G = 0.283$) were significant, as was the *testing context* × *stimulus novelty* interaction ($F_{(1,11)} = 46.4$, $p < 0.0001$, $\eta^2_G = 0.317$). Paired t tests showed that accuracy increased more to the trained than untrained targets that were presented alone (i.e., hT_F-alone > vT_N-alone; $M_D = 0.25$, $t_{11} = 6.25$, $p < 0.0001$) but not when they were embedded in a non-informative context ($hT_F + vC \not> vT_N + hC$; $M_D = -0.01$, $t_{11} = -0.42$, $p = 0.686$). Hence, as was found in Experiments 1 and 2, training with (hT-alone textures resulted in stimulus specific learning ($hT_F > vT_N$) that did not generalize to a condition in which the familiar targets were embedded in a new, non-informative context ($hT_F + vC \not> vT_N + hC$).

In the target+context group (grey bars in Figure 5.4B), the intercept was significantly different from zero ($F_{(1,11)} = 45.1$, $p < 0.0001$, $\eta^2_G = 0.404$), indicating that overall, accuracy was higher after training. The main effect of *testing context* ($F_{(1,11)} = 0.724$, $p = 0.413$, $\eta^2_G = 0.009$) was not significant, but the main effect of *stimulus novelty* ($F_{(1,11)} = 5.90$, $p = 0.033$, $\eta^2_G = 0.194$) was significant, indicating

that accuracy improved more for the familiar than unfamiliar targets. Interestingly, unlike the target-alone group, the *testing context* × *stimulus novelty* interaction ($F_{(1,11)} = 1.79$, $p = 0.208$, $\eta^2_G = 0.039$) was not significant. Hence, the ANOVA suggests that the orientation-specific improvement in identification accuracy was similar for targets embedded within an informative context or presented alone, despite the fact that participants were trained to identify targets only within a non-informative context. For example, inspection of the two left grey bars in Figure 5.4B indicates that there was a significant increase in response accuracy for the familiar horizontal targets presented alone (hT_F-alone, $M = 0.110$, $t_{11} = 5.04$, $p = 0.0004$) and embedded in a context ($hT_F + vC$, $M = 0.093$, $t_{11} = 3.53$, $p = 0.0047$), and these improvements did not differ from each other ($M_D = -0.017$, $t_{11} = -0.691$, $p = 0.504$). The change in response accuracy was not significant for novel vertical targets presented alone (vT_N-alone, $M = 0.003$, $t_{11} = 0.109$, $p = 0.915$), but was significant for novel vertical targets embedded in a non-informative context ($vT_N + hC$, $M = 0.0486$, $t_{11} = 2.61$, $p = 0.0243$), although the change in the two conditions did not differ from each other ($M_D = 0.0458$, $t_{11} = 1.38$, $p = 0.194$). Paired t tests failed to find strong evidence of stimulus-specific learning ($hT_F + vC \not> vT_N + hC$; $M_D = 0.0444$, $t_{11} = 1.13$, $p = 0.281$), but did find evidence for generalization of learning to a condition in which the familiar targets were no longer embedded in a non-informative context (hT_F-alone > vT_N-alone; $M_D = 0.107$, $t_{11} = 2.76$, $p = 0.0185$). Unlike what was found in Experiments 1 and 2, and the target-alone group in Experiment 3, training with $hT + vC$ textures resulted in learning that generalized to the trained targets presented alone, despite being trained with the targets embedded in a non-informative context.

The ANOVAs on the pre- and post-testing accuracy data suggest that the target+context group demonstrated learning that was less stimulus-specific than learning the target-alone group, but also demonstrated greater generalization of learning to familiar targets when the context was removed. Other comparisons confirm this result: stimulus-specific learning, which was defined for the target-alone group as hT_F-alone > vT_N-alone, and for the target+context group as $hT_F + vC > vT_N + hC$, was significantly greater in the target-alone group ($M_D = 0.201$, $t_{22} = 3.63$, $p = 0.0015$). Context-generalization, which was defined for the target-alone group as $hT_F + vC > vT_N + hC$, and for the target+context group as hT_F-alone > vT_N-alone, was greater in the target+context group ($M_D = -0.117$, $t_{22} = -2.58$, $p = 0.0172$). These results differ notably from those obtained in Experiments 1 and 2, where we found no evidence

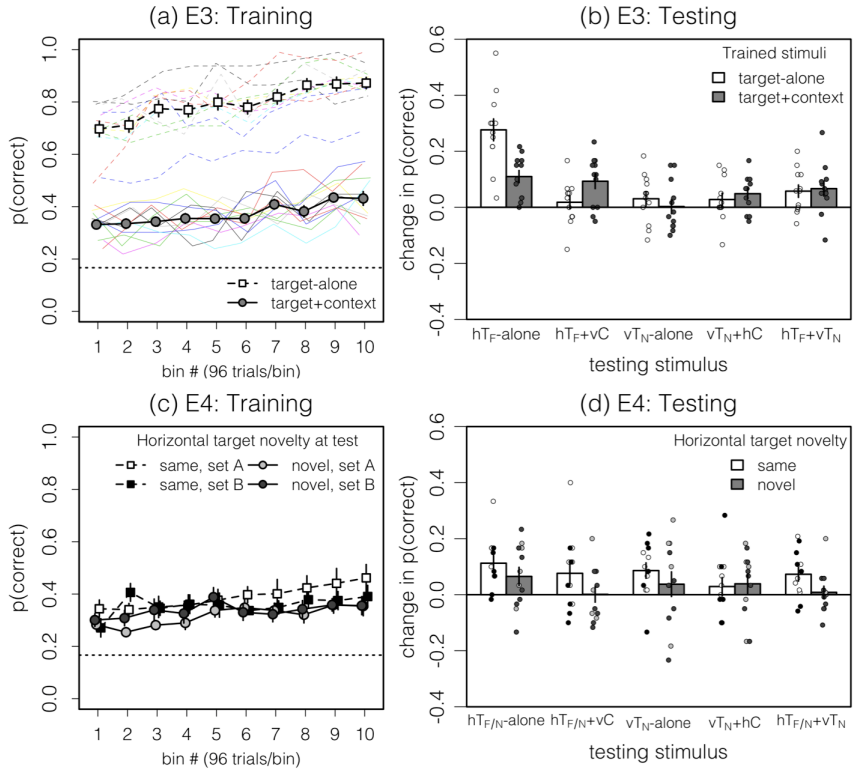

Figure 5.4 – Results from Experiments 3 and 4. Refer to Figure 5.2 caption for details of figure. **A)** Training data for Experiment 3: Performance was significantly worse for the target+context trained group (grey) than the target-alone trained group (white) throughout all of training. The learning rate also was lower for the target+context training group. **B)** Testing data for Experiment 3: Training with target-alone textures (white) led to the greatest improvements to the hT_F-alone textures, whereas training with target+context textures (grey) led to context-generalization: the greatest improvements were seen in both the hT_F+vC and hT_F-alone textures. **C)** Training data for Experiment 4: Groups were trained with different sets of target+context textures, but performed similarly throughout training trials. **D)** Testing data for Experiment 4: Training to identify horizontal targets within a vertical context did not meaningfully transfer to horizontal but novel targets (grey $\mathrm{hT_{F/N}}$ bars) relative to horizontal but familiar targets (white $\mathrm{hT_{F/N}}$ bars). Individual subject data points are consistent with the colour coding used in C.

of context-generalization in either training condition.

Lastly, we compared the effect of training on response accuracy for $hT_F + vT_N$ textures with hT_F-alone and $hT_F + vT_N$ textures. Similar to what was found in Experiment 1, the effect of training on response accuracy in the target-alone trained group was smaller for $hT_F + vT_N$ textures than hT_F-alone textures ($M_D = -0.219$, $t_{11} = -5.38$, $p = 0.0002$), and slightly but not significantly higher than $hT_F + vC$ textures ($M_D = 0.040$, $t_{11} = 1.94$, $p = 0.0789$). Therefore, training in the target-alone condition had similar effects on accuracy for test stimuli in which the vertical context was or was not informative. In the target+context trained group, the effect of training on accuracy for the $hT_F + vT_N$ textures was not significantly different than than the effect of training on $hT_F + vC$ textures ($M_D = -0.026$, $t_{11} = -0.946$, $p = 0.365$), and was slightly but not significantly lower than the effect on hT_F-alone textures ($M_D = -0.043$, $t_{11} = -2.17$, $p = 0.0527$). Unlike Experiments 1 and 2, and the target-alone group in the current experiment, training with target+context textures led to similar effects of training for textures with non-informative or informative vertical structures. Although different from Experiments 1 and 2, this is not surprising given the stimulus: here, the context in target+context textures was different on every trial, and therefore, to the naïve participant, the informative vertical structure in $hT_F + vT_N$ textures is no different than the uninformative vertical structure in $hT_F + vC$ textures.

5.5.3 Discussion

The aim of Experiment 3 was to test the hypothesis that randomly varying the non-informative context would discourage participants from including the context in their learned representation of the targets. Our results are consistent with this hypothesis. Training to identify targets embedded in a novel context on every trial resulted in increased accuracy for the trained targets embedded in a non-diagnostic context, and, critically, to the trained targets presented alone. This is our first demonstration of context-generalization. However, acquiring context generalizable learning came at the cost of completing a significantly more difficult training task: overall accuracy, and the rate of improvement during training, was significantly reduced.

Our demonstration that individuals learn to identify a fixed set of horizontal targets in a virtually infinite set of vertical contexts, and then successfully identify those targets when the context is removed, is consistent with the idea that randomization of the non-informative context made it easier for individuals to selectively re-weight

orientation-selective channels to produce a decision variable that was influenced more by horizontal structure than vertical structure. Such re-weighting might occur by reducing the weight of vertically-tuned channels; however, some of our results are inconsistent with this idea. For example, response accuracy during the training phase was much lower in the target+context group, and the difference between groups increased with training. Hence, it seems clear that observers could not ignore the non-informative, vertical context. Furthermore, if learning reflected the down-weighting of vertically tuned channels, then post-training accuracy for textures with novel, vertically oriented targets (i.e., vT-alone & $vT + hC$ textures) should have decreased relative to pre-training. We do not see this effect: Accuracy in those conditions (i.e., the third and forth grey bars in Figure 5.4B) did not decrease, which suggests that training did not cause participants to discount the responses of vertically tuned channels. Instead, our results suggest that training resulted in an increase in the weight assigned to horizontally tuned channels.

An important result found in the first three experiments is that training with target-alone textures did not generalize to target+context textures. This finding, in conjunction with the context-generalization observed for the target+context trained group, suggests that the re-weighting of orientation channels may require a non-uniform distribution of diagnostic value in the visual signal. Restricting the signal to a certain orientation bandwidth is not adequate to force the up-weighting of those orientation channels, but rather, certain channels must contain a diagnostic signal (e.g., horizontal target), and at least *some* other channels must contain a non-diagnostic signal (e.g., vertical context). It is with a non-uniform distribution of diagnostic information the decision variable will differentially rely on the signal from certain orientations. If, instead, there was a non-uniform distribution of the visual *signal* (e.g., hT-alone textures), then the decision variable would not need to differentially rely on certain orientations for the diagnostic information, since the incoming signal would already be entirely diagnostic. In other words, the fact that we do not see context-generalization when training on target-alone textures, but do see context-generalization when training on target+context textures, suggests that channel re-weighting requires some sort of selective attention that extracts the diagnostic signal that is embedded in non-diagnostic signals. For instance, learning to identify target+context textures may be best achieved through the selective attention to, or the increased processing of, the horizontally oriented target structure embedded within non-informative structure

in all other orientations. However, learning to identify target-alone textures does not demand selective attention/processing, as the stimulus itself is limited in what it contains, so no additional processing/perceptual strategy is needed to separate target from context. With target-alone textures, the target's informative signal is delivered to the decision variable as a pure representation without the need to filter out any non-informative contextual signals.

If learning is happening by selective re-weighting of horizontal orientation channels, then it remains unclear if the re-weighting of channels reflects a refinement (or creation of) perceptual templates specific to the six trained horizontal targets, or if it reflects the refinement of all, or most, horizontally tuned channels that will therefore allow better identification of any horizontally oriented target. The stimulus-specificity often observed in perceptual learning suggests that we should not expect transfer of learning to novel horizontally oriented targets, but some evidence suggests that training with a new set of textures on every trial is not only slow like the target+context condition, but also produces learning that transfers to novel textures (Hussain et al., 2012a). We test for the transfer to novel, horizontal targets in the Experiment 4.

Regardless of the mechanism, the evidence of context-generalization suggests that randomly varying non-informative context altered what was learned during training, so that visual processing more effectively ignored the non-informative vertical structure. Alternatively, exposure to a randomly varying context may simply have slowed the acquisition of stimulus-specific learning (Jeter et al., 2010). The change in response accuracy during training was significantly less in the target+context group than the target-alone group. Therefore, if stimulus-specificity emerges gradually over the course of training, then perhaps differences between the two groups would diminish if the rate of learning was equated across conditions. We examine this possibility in Experiment 5.

5.6 Experiment 4: Target- and orientation-specificity of learning

Experiment 3 demonstrated that training to identify horizontal targets embedded in a context produces learning that transfers to those same targets presented alone. In Experiment 4, we examine whether such learning also transfers to novel horizontally-

oriented targets. Before and after training on a set of six target+context textures, participants were either tested on trained (familiar) or untrained (novel) horizontal targets, presented alone and embedded in a vertical context. Like previous experiments, they were also tested on vertical targets presented alone or embedded in a horizontal context. The condition where participants viewed familiar horizontal targets at test is a replication of the target+context condition from Experiment 3.

5.6.1 Methods

Horizontal targets were derived from two sets of noise patterns: either the same noise patterns used in deriving horizontal targets in Experiment 3 (set A), or a novel set of patterns (set B). Half of all participants were trained on set A, and half were trained on set B. In each group, half of the participants were tested with the same set they were trained on ("familiar" group), and half were tested with the set they were not trained on ("novel" group). During testing, as with previous experiments, the vertical targets always were novel textures. All groups were trained with target+context ($hT + vC$) textures. All other aspects of the stimuli, apparatus, and experimental procedure were the same as in Experiment 3.

Experiment 3 found that the rate of learning and overall accuracy during training were low when the non-informative context varied randomly across trials. To maintain the same exclusion criteria used in the previous experiments (i.e., improvements in accuracy during training was greater than 0) and have our desired number of participants per training group, we excluded 19 out of the 43 recruited participants[1]. The final sample consisted of 24 naïve participants (5 males, range = 18-22 years old, M = 19.2, SD = 1.41), 12 in each training group.

5.6.2 Results

5.6.2.1 Training Phase

Response accuracy during the training phase is plotted in Figure 5.4C. Our analysis of training data was focused on ensuring the two groups (familiar vs. novel horizontal targets at test) improved similarly during training, so the data was submitted to a 2 (*horizontal novelty*) × 2 (*target set*) factorial ANOVA. The ANOVA on overall

[1]We address the high exclusion rate using an alternative paradigm in Experiment 6.

response accuracy found that the main effects of *novelty group* ($F_{(1,20)} = 2.60$, $p = 0.122$, $\eta^2_G = 0.115$) and *target set* ($F_{(1,20)} = 0.040$, $p = 0.843$, $\eta^2_G = 0.002$), and the interaction ($F_{(1,20)} = 0.849$, $p = 0.368$, $\eta^2_G = 0.041$) were not significant. Furthermore, the ANOVA on the linear trend (of accuracy across training bins) scores found that the grand mean of the linear trend scores differed significantly from zero ($F_{(1,20)} = 30.0$, $p < 0.0001$, $\eta^2_G = 0.600$), and that the linear trend did not differ significantly across groups ($F_{(1,20)} = 0.485$, $p = 0.494$, $\eta^2_G = 0.024$), which suggests that accuracy did improve with training and that the rate of learning, as indexed by the linear trend, did not differ between the novelty groups. The linear trend did differ significantly between groups trained with the two sets of targets ($F_{(1,20)} = 4.51$, $p = 0.0464$, $\eta^2_G = 0.184$), reflecting a smaller linear trend in the group who saw target set B ($M = 0.048$, $t_{12} = 2.83$, $p = 0.0152$) than target set A ($M = 0.112$, $t_{10} = 4.90$, $p = 0.0006$); however, the effect of target set did not interact with novelty group ($F_{(1,20)} = 0.140$, $p = 0.712$, $\eta^2_G = 007$). These analyses suggest that the rate of learning during training differed between participants trained with stimulus sets A and B, but not between subjects who were tested with novel or familiar textures during testing.

5.6.2.2 Pre- & Post-training Test Phases

The change in accuracy from pre- to post-training is shown for each group and test condition in Figure 5.4D. Inspection of the individual data points in Figure 5.4D suggests that there were no obvious differences during the test phase between participants trained with stimulus sets A and B. The ANOVA on the difference scores revealed that, on average, accuracy was higher after training ($F_{(1,22)} = 15.9$, $p = 0.0006$, $\eta^2_G = 0.201$). The main effects of *training group* ($F_{(1,22)} = 2.06$, $p = 0.165$, $\eta^2_G = 0.032$), *testing context* ($F_{(1,22)} = 2.99$, $p = 0.0979$, $\eta^2_G = 0.030$), and *stimulus novelty* ($F_{(1,22)} = 0.709$, $p = 0.409$, $\eta^2_G = 0.005$) were not significant. Furthermore, the *training group* × *stimulus novelty* interaction was not significant ($F_{(1,22)} = 1.21$, $p = 0.284$, $\eta^2_G = 0.009$), nor were all of the other two- and three-way interactions ($F < 1$, $p > 0.38$, $\eta^2_G < 0.01$, in all cases). Hence, the ANOVA on changes in pre- and post-training accuracy suggests that overall accuracy was better after training ($M = 0.056$), but that the change in accuracy did not differ significantly between training groups or stimulus conditions. This pattern may be interpreted as evidence of generalization to stimulus conditions, but, given that the ANOVA suggests there were no differences in improvement of hT_F-alone and $hT_F + vC$ textures from vT_N-alone

and $vT_N + hC$ textures, which were included as control stimuli, we interpret the lack of significant effects in the ANOVA as evidence that stimulus-related learning was minimal.

To directly compare context generalization across Experiments 3 and 4, we analyzed the data from the familiar training group in Experiment 4 and the data from the target+context training group in Experiment 3 in a 2 (*testing context*) × 2 (*stimulus novelty*) × 2 (*experiment*) ANOVA. Note that in Experiment 3, evidence for context generalization was provided by a significant main effect of *stimulus novelty* that *did not* interact with *testing context*. The ANOVA here also revealed a significant main effect of *stimulus novelty* ($F_{(1,22)} = 6.32$, $p = 0.0197$, $\eta^2_G = 0.084$) that did not interact with *testing context* ($F_{(1,22)} = 0.305$, $p = 0.586$, $\eta^2_G = 0.003$) or *experiment* ($F_{(1,22)} = 0.755$, $p = 0.394$, $\eta^2_G = 0.011$). The main effects of *testing context* and *experiment* were not significant ($F < 1$, $p > 0.38$, $\eta^2_G < 0.008$, in each case), nor was their two-way interaction ($F_{(1,22)} = 2.90$, $p = 0.103$, $\eta^2_G = 0.026$), or their three-way interaction with *stimulus novelty* ($F_{(1,22)} = 1.22$, $p = 0.281$, $\eta^2_G = 0.012$). Hence, the evidence for context generalization was comparable and consistent across Experiments 3 and 4.

5.6.3 Discussion

Experiment 4 examined whether the context-generalizable learning we observed in Experiment 3 reflected participants learning to discriminate a specific set of six horizontally oriented targets in any context, or whether they were learning to better extract horizontal structure from textures, and therefore better identify any horizontally oriented targets in any context. Results from Experiment 4 were not conclusive, but given the similarities with results from Experiment 3, we combined Experiments 3 and 4 for a more powerful analysis. The combined results suggest that training to identify horizontal targets embedded within a vertical context produces learning that generalizes to familiar/trained targets in the absence of a vertical context, but does not transfer to identifying novel/untrained horizontal targets.

The current experiment failed to find statistically significant evidence for transfer of learning to novel, horizontal targets. Such evidence would manifest as greater improvements for hT_N or $hT_N + vC$ relative to vT_N-alone or $vT_N + hC$ stimuli, respectively; this difference would first appear as a main effect of *stimulus novelty*, but it was not significant. However, inspection of Figure 5.4D reveals that participants improved more, on average, at identifying novel horizontal targets when they were presented

alone (hT_N-alone, first left grey bar) than novel horizontal targets presented within a context ($hT_N + vC$, second left grey bar). Hence, it is still possible that participants are able to better identify novel horizontal targets, but given small learning effects, it was too difficult to obtain significant effects. In fact, for us to detect a significant effect of context-generalization to familiar targets, we had to combine data from Experiments 3 and 4. Therefore, it is possible that generalization to novel targets occurred, but we lacked statistical power to detect it. In the upcoming set of studies, we redesign our experiment to be more sensitive to small effects (Experiment 5), and increase the reliability of the effect (Experiment 6).

5.7 Experiment 5: The role of difficulty during training on the context-generalization of learning

Experiment 3 revealed that participants who were trained with target+context stimuli exhibited significantly lower identification accuracy and a slower rate of improvement compared to participants trained with target-alone stimuli. These differences in accuracy may have produced a slower acquisition of stimulus-specific learning (Jeter et al., 2010) in the target+context trained group than the target-alone trained group. The current experiment examines this possibility by examining the effects of increasing the difficulty of the task with low-contrast, target-alone textures. In addition, the current experiment re-examines how learning transfers to novel targets in the trained orientation. Experiment 4 indicated that some transfer of learning to novel targets may be possible, but context generalization and transfer to novel targets were not measured in the same participants. Therefore, we modified the experimental design such that stimulus-specificity, context-generalization, and transfer to novel horizontal targets were all within-subject variables.

5.7.1 Methods

Thirty-seven naïve participants completed the task, but 13 were excluded using the same exclusion criteria as Experiment 3 (i.e., they failed to show any improvement in response accuracy during training). The final sample comprised 24 participants (8

males, range = 17-35 years old, M = 21.2, SD = 5.1). The participants were assigned randomly to two groups of 12 individuals who were trained with either target-alone (hT-alone) or target+context ($hT + vC$) textures. All participants completed the same pre- and post-training testing conditions: i) familiar horizontal targets presented alone (hT_F-alone); ii) familiar horizontal targets presented within a non-informative vertical context ($hT_F + vC$); iii) novel horizontal targets presented alone (hT_N-alone); iv) novel horizontal targets presented within a non-informative vertical context ($hT_N + vC$); v) novel vertical targets presented alone (vT_N-alone); and vi) novel vertical targets presented within a non-informative horizontal context ($vT_N + hC$).

To equate performance in the two training groups, the contrast of the target texture was lower in the target-alone stimulus than in the target+context stimulus. A pilot experiment on three naïve participants, not included in the final sample, indicated that response accuracy was approximately equal when target RMS contrast was 0.0088 and 0.035 in the target-alone and target+context conditions, respectively. Therefore, we used these values of RMS contrast for the training and testing components of the main experiment. All other aspects of the stimuli, apparatus, and experimental procedure were the same as in Experiments 3 and 4.

5.7.2 Results

5.7.2.1 Training Phase

Performance during training is plotted in Figure 5.5A. Response accuracy during training was, on average, not significantly different between the two training groups ($M_D = 0.04$, $t_{22} = 1.62$, $p = 0.120$). The linear trend of accuracy across training bins was significant ($F_{(1,22)} = 56.2$, $p < 0.0001$, $\eta^2_G = 0.719$), indicating that accuracy improved with training, but the linear trend differed significantly between training groups ($F_{(1,22)} = 6.84$, $p = 0.0158$, $\eta^2_G = 0.237$). Follow-up tests indicated that the linear trend was significant in both the target-alone group ($M = 0.137$, $t_{11} = 5.58$, $p < 0.0001$) and the target+context group ($M = 0.066$, $t_{11} = 5.75$, $p < 0.0001$), but it was smaller in the target+context group. These analyses suggest that overall accuracy was similar in the two groups but, as was found in Experiment 3, accuracy improved at a faster rate in the target-alone group. Figure 5.5A indicates that target-alone trained participants began to outperform the target+context group on bin 6, which is the start of day 2 training, hinting at a potential role of consolidation during sleep in the

target-alone group (Censor et al., 2006). Despite these group differences, performance during training was, overall, similar between the groups, and our manipulation of target contrast dramatically reduced the group differences found in the previous experiments.

5.7.2.2 Pre- & Post-training Test Phases

Changes in response accuracy from pre- to post-training are shown in Figure 5.5B. Difference scores were submitted to a 2 (*training condition*) × 2 (*testing context*) × 3 (*stimulus type*) ANOVA. We grouped familiar horizontal targets (hT_F), novel horizontal targets (hT_N), and novel vertical targets (vT_N) into a single factor of *stimulus type* because familiarity and orientation were not factorially crossed. The ANOVA revealed that overall, response accuracy increased significantly ($F_{(1,22)} = 4.98$, $p = 0.0361$, $\eta^2_G = 0.061$). The main effects of *training condition* ($F_{(1,22)} = 0.0005$, $p = 0.982$, $\eta^2_G < 0.001$), *testing context* ($F_{(1,22)} = 0.907$, $p = 0.351$, $\eta^2_G = 0.007$), and *stimulus type* ($F_{(2,44)} = 1.20$, $p = 0.310$, $\eta^2_G = 0.014$) were not significant. The two-way interactions of *training condition* × *stimulus type* ($F_{(2,44)} = 1.09$, $p = 0.346$, $\eta^2_G = 0.013$), *stimulus type* × *testing context* ($F_{(2,44)} = 1.66$, $p = 0.203$, $\eta^2_G = 0.020$), and *training condition* × *testing context* ($F_{(1,22)} = 4.22$, $p = 0.0521$, $\eta^2_G = 0.032$) were not significant. Finally, the three-way interaction also was not significant ($F_{(2,44)} = 0.167$, $p = 0.847$, $\eta^2_G = 0.002$). In summary, the ANOVA showed that accuracy improved after training, but that the improvement did not vary significantly across groups or conditions. Separate 2 (testing context) × 3 (stimulus type) ANOVAs performed on the data from each testing group also failed to find significant differences across stimulus conditions.

The pattern of results in Experiments 3 (Figure 5.4B), 4 (white bars in Figure 5.4D), and 5 (Figure 5.5B) look similar: In each experiment, the target-alone group showed an effect of training to hT_F-alone but not $hT_F + vC$ textures, but the target+context trained group showed similar improvements to $hT_F + vC$ and hT_F-alone textures. Hence, we combined the data across the Experiments 3 and 5 for the target-alone training condition, and Experiments 3, 4, and 5 for the familiar target+context training condition, and assessed stimulus-specificity and context-generalization using the larger, combined samples. For the target-alone group, stimulus-specificity was defined as the difference in improvements between hT_F-alone and vT_N-alone stimuli, and context-generalization was defined as the difference in improvements between $hT_F + vC$ and $vT_N + hC$ stimuli. For the target+context group, stimulus-specificity was defined as the difference in improvements between $hT_F + vC$ and $vT_N + hC$ stimuli, and context

generalization was defined as the difference in improvements between hT_F-alone and vT_N-alone stimuli. We found that for the target-alone group, stimulus-specific learning (hT_F-alone $> vT_N$-alone; $M_D = 0.137$, $t_{23} = 3.72$, $p = 0.0011$) was significant, but context-generalization ($hT_F + vC \not> vT_N + hC$; $M_D = -0.015$, $t_{23} = -0.879$, $p = 0.388$) was not significant. Opposite results were obtained for the target+context group: stimulus-specific learning ($hT_F + vC \not> vT_N + hC$; $M_D = 0.042$, $t_{35} = 1.81$, $p = 0.0796$) was not significant, but context-generalization (hT_F-alone $> vT_N$-alone; $M_D = 0.064$, $t_{35} = 2.91$, $p = 0.0062$) was significant. Stimulus-specific learning in the target-alone group was significantly greater than the target+context group ($M_D = 0.096$, $t_{58} = 2.32$,$p = 0.0239$), and context generalization in the target+context group was significantly greater than the target-alone group ($M_D = 0.080$, $t_{58} = 2.60$,$p = 0.0117$). Lastly, we combined the data across experiments 4 and 5 for the novel target+context condition, and assessed generalization to novel targets presented alone or embedded within an uninformative context. We found that generalization to novel targets presented in a non-informative vertical context ($hT_N + vC \not> vT_N + hC$; $M_D = -0.026$, $t_{23} = -1.15$, $p = 0.263$ was not significant, and generalization to novel targets presented alone (hT_N-alone $\not> vT_N$-alone; $M_D = 0.015$, $t_{23} = 0.565$, $p = 0.577$ also was not significant. Therefore, after achieving greater power by combining participants from Experiments 4 and 5, we see that there was no significant effect of training on generalization to novel but horizontally oriented targets.

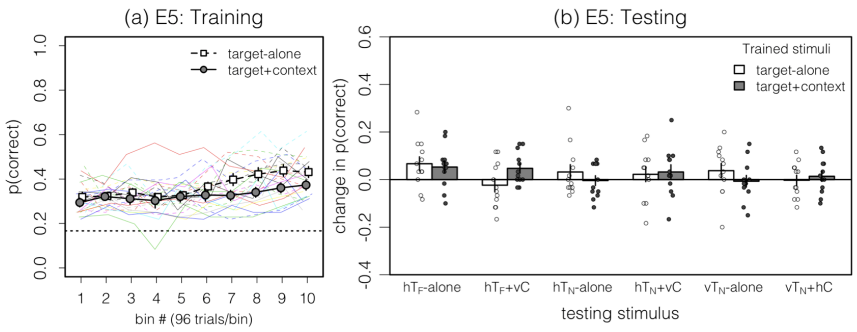

Figure 5.5 – Results from Experiment 5. Refer to Figure 5.2 for details of figure. **A)** Training accuracy for participants trained on either target-alone (white) or target+context (grey) textures. Target-alone textures had lower contrast to match the difficulty in identifying target+context textures. **B)** Training to identify low contrast target-alone textures still produced specific learning only to the hT_F-alone textures. Training to identify target+context textures produced context generalizable learning: improvements were greatest for both $hT_F + vC$ and hT_F-alone textures. In both training groups, there were no meaningful changes in response accuracy to hT_N-alone, $hT_N + vC$, vT_N-alone, and $vT_N + hC$ textures.

5.7.3 Discussion

Although the effects in Experiment 5 were not statistically significant, the pattern was consistent with Experiments 3 and 4: the effect of training with target+context textures generalizes to better identification accuracy of familiar targets presented alone, but training with target-alone textures, even when they are approximately difficulty-matched with target+context textures, does not produce learning that generalizes to the same targets in the presence of a context. Given that the effects are relatively small, we rely on the analyses on the data combined across experiments for these conclusions. Furthermore, although there was no statistically significant evidence, the trend of the means is consistent with the notion that the context-generalizable learning from target+context training does not reflect a broader, orientation-specific learning, as there is no indication that performance to novel targets in the trained orientation, either with or without a context, improved after training. Also, an important factor that likely accounts for the very small changes in response accuracy is that during testing, for all training groups, the target-alone textures were presented with reduced contrast. Therefore, target+context trained participants, who were seeing the target (and the context) at relatively high contrasts during the training phase, were being asked to identify those targets in isolation at much lower contrasts during the test

phase. Nonetheless, the direction of the non-significant effects are all consistent with findings from earlier experiments.

Despite the consistent *patterns* of results in Experiments 3, 4 and 5, the lack of significant effects is not enough to make conclusions about the context generalizability of the learning, or its lack of transfer to novel targets. It is necessary to know if, given a more effective training paradigm, participants show similar patterns of context-generalization and target specificity. Therefore, the following experiment examined whether a more effective training paradigm affects the characteristics of learning.

5.8 Experiment 6: Improving the training paradigm to promote learning

Experiments 3 to 5 have found a consistent but very small effect of context-generalization following training with targets embedded in a context. One criticism of those experiments is that response accuracy improved only slightly during training, even after excluding participants who failed to exhibit any learning. Therefore, our next experiment used a training paradigm that we expected would increase the reliability and magnitude of learning in our task.

To better discriminate targets in the target+context condition, participants have to first distinguish what the horizontal structure of the texture is. Therefore, the current experiment manipulated the stimuli to help participants distinguish diagnostic and non-diagnostic structure. Specifically, participants were trained on target+context textures where the contrast of the non-diagnostic context was varied using a one-up-one-down staircase while the contrast of the target was held constant. As participants responded correctly, the contrast of the context was increased (making target identification harder); as they responded incorrectly, the contrast was decreased (making target identification easier). Hence, our method essentially measured how much non-informative context an individual could tolerate while identifying the targets.

5.8.1 Methods

Twelve naïve participants (1 male, range = 17-23 years old, M = 18.75, SD = 1.60) participated in this experiment. All twelve participants were included in the analysis as they all exhibited some form of performance improvements during training, as

described in the results. Throughout training, all 12 participants saw target+context ($hT + vC$) textures with a fixed target RMS contrast of 0.035, and a context RMS contrast that was varied across trials using a 1-up-1-down staircase. The staircase could vary RMS contrast from 0.0035 to 0.35 in steps of 0.05 log units. The six items on the response screen had target contrasts of 0.035 and context contrasts that were the same as the context contrast that was presented on that trial. Two training sessions were completed on consecutive days; each session contained two blocks, and in each block the context contrast was varied with three interleaved staircases that terminated after 160 trials. Staircases always started from the lowest context contrast to ensure participants adequately experienced the transformation of target-alone to target+context, since at the lowest contrast, the context was barely visible relative to the target. The RMS contrast that corresponded to 50% correct (Levitt, 1971) was estimated by first averaging the last four reversals of each staircase, and then averaging the values across the three staircases in each block, resulting in four estimates of threshold per participant. Note that *higher* thresholds correspond to better performance in the form of increased tolerance of contextual structure while identifying target+context textures.

The staircase was used exclusively during training trials. In all testing conditions, the target, and when present, the context, were each presented with a fixed, suprathreshold contrast of 0.035. All other aspects of the stimuli, apparatus, and experimental procedure were the same as in Experiment 5.

5.8.2 Results

5.8.2.1 Training Phase

Contrast thresholds obtained from the training phase are plotted in Figure 5.6A. Contrast thresholds were submitted to a one-way repeated measures ANOVA with block as a four-level factor. The ANOVA revealed that contrast thresholds varied significantly across blocks ($F_{(3,33)} = 11.9$, $p_{HF} < 0.0001$, $\eta^2_G = 0.261$). Follow-up focussed contrasts tested for within-session learning (weights: [-1 1 -1 1] for blocks 1 to 4, respectively) and between session learning (weights: [-1 -1 1 1] for blocks 1 to 4, respectively). Scores for within-session learning were greater than zero for 8/12 participants, and scores for between-session learning were greater than zero for 11/12 participants. We included all participants in our analysis because everyone had at least

one positive within-session or between-session learning score. The focussed contrasts revealed that changes within-session ($M = 0.0013$, $t_{11} = 0.475$, $p = 0.644$) were not significantly different from zero, but that changes between-sessions ($M = 0.018$, $t_{11} = 5.20$, $p = 0.0003$) were significant. These results are consistent with the idea that performance improved across days but not within sessions, perhaps reflecting the effects of sleep consolidation (Censor et al., 2006). Additionally, the linear contrast scores for between-session learning for 11 out of 12 observers were positive (range=0.009-0.042, M=0.021, SD=0.010), with only one participant having a slightly negative score (-0.005). We confirmed that thresholds did not significantly differ between blocks 1 and 2 (day 1: $M_D = -0.00008$, $t_{11} = -0.0576$, $p = 0.955$) or between blocks 3 and 4 (day 2: $M_D = 0.001$, $t_{11} = 0.558$, $p = 0.588$), but they did significantly differ between blocks 2 and 3 (across days 1 and 2: $M_D = 0.0085$, $t_{11} = 3.98$, $p = 0.0022$). The analysis of the training data suggests that participants demonstrated a significant amount of learning. On average, contrast thresholds increased from 0.018 to 0.028, indicating that participants were able to correctly identify 50% of trials where the contrast of the context was approximately 80% of the contrast of the target (0.035).

To better relate results from the staircase paradigm to our previous experiments, we plotted the average proportion correct at each context contrast level, collapsed across all staircases in a single day, in Figure 5.6B. The dotted vertical line in Figure 5.6B represents the point on the x-axis where the context contrast equalled the target contrast, as in Experiments 2-5. The horizontal lines indicate the average accuracy to textures with equal context and target contrast on day 1 (solid) and day 2 (dashed). The data reveals that the average improvement in accuracy from day 1 ($M = 0.305$) to day 2 ($M = 0.425$) was 0.12, which is consistent with the average improvements in Experiments 3-5. Furthermore, improvements appear to be approximately equal across all contrast levels and were not restricted to the lower contrast context conditions which they were exposed to for more trials.

5.8.2.2 Pre- & Post-training Test Phases

The change in accuracy from pre- to post-training in the testing conditions is plotted in Figure 5.6C. The 2 (*testing context*) × 3 (*stimulus type*) ANOVA revealed that, on average, difference scores differed from zero ($F_{(1,11)} = 6.81$, $p = 0.0242$, $\eta^2_G = 0.125$). The main effect of *stimulus type* ($F_{(2,22)} = 7.07$, $p = 0.0043$, $\eta^2_G = 0.140$) was significant, but the main effect of *testing context* ($F_{(1,11)} = 1.23$, $p = 0.291$, $\eta^2_G = 0.025$) and the

stimulus type × *testing context* interaction ($F_{(2,22)} = 1.10$, $p = 0.351$, $\eta^2_G = 0.028$) were not. The analysis suggests that there was a difference between familiar horizontal, novel horizontal, and novel vertical targets, but that effect was not modulated significantly by the presence or absence of a vertical context. Follow-up t tests revealed that the main effect of *stimulus type* was due to significantly greater increases in accuracy for familiar horizontal targets than novel horizontal targets ($M_D = 0.083$, $t_{11} = 2.55$, $p = 0.0268$) and novel vertical targets ($M_D = 0.1000$, $t_{11} = 4.15$, $p = 0.0016$). The changes in accuracy for novel horizontal and novel vertical targets were not significantly different ($M_D = 0.017$, $t_{11} = 0.593$, $p = 0.565$). Finally, one-sample t tests revealed that improvements were significantly different from zero for the familiar horizontal targets ($M = 0.102$, $t_{11} = 4.86$, $p = 0.0005$), but not for the novel horizontal ($M = 0.0187$, $t_{11} = 0.674$, $p = 0.514$) or novel vertical targets ($M = 0.0021$, $t_{11} = 0.114$, $p = 0.911$).

The results indicate that training to identify horizontal targets embedded in a vertical context, with the contrast of the context increasing using a staircase, lead to improved performance during the training phase in all 12 participants. The learning transfers to the same targets presented without any context (i.e., context-generalization), but it does not transfer to novel horizontal targets or novel vertical targets, regardless of context. These results are qualitatively similar to the results from Experiments 3-5, with the advantage that no participants were excluded, attesting to the effectiveness of the training paradigm in promoting learning. To quantitatively investigate differences across experiments, we evaluated linear comparisons between Experiment 6 and the target+context groups from Experiments 3-5. The effect of training (i.e., the difference between pre- and post-training accuracy) in Experiment 6 (weight = 1) did not differ from the training effect in Experiments 3-5 (weights = -1/3 each) for familiar horizontal targets embedded within a vertical context ($hT_F + vC$: $\psi = 0.010$, $t_{44} = 0.271$, $p = 0.788$), familiar horizontal targets presented alone (hT_F-alone: $\psi = 0.031$, $t_{44} = 0.940$, $p = 0.353$), or for novel vertical targets presented alone (vT_N-alone: $\psi = 0.013$, $t_{44} = 0.392$, $p = 0.697$). However, the effect of training was smaller in Experiment 6 than Experiments 3-5 for novel vertical targets that were presented within a horizontal context ($vT_N + hC$: $\psi = -0.067$, $t_{44} = -2.52$, $p = 0.016$). Finally, the effect of training in Experiment 6 (weight = 1) did not differ from Experiments 4-5 (weights = -1/2 each) for novel horizontal targets embedded within a vertical context ($hT_N + vC$: $\psi = 0.008$, $t_{33} = 0.241$, $p = 0.811$) or presented alone (hT_F-alone: $\psi = -0.018$, $t_{33} = 0.462$, $p = 0.647$). These results further suggest that the staircase

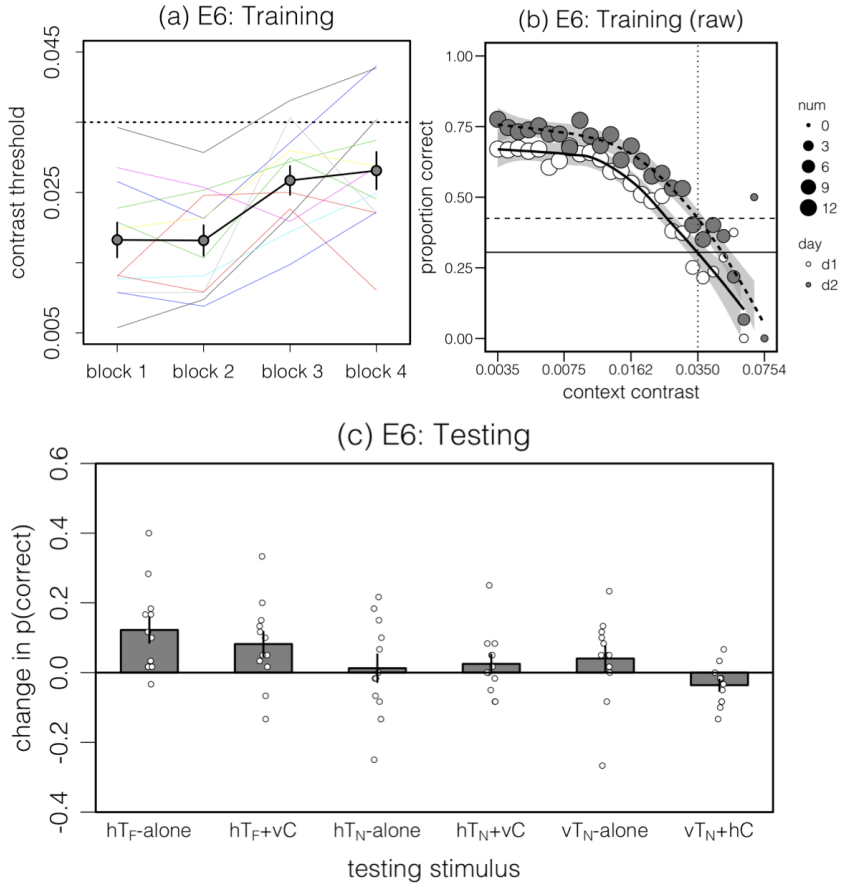

Figure 5.6 – Results from Experiment 6. **A)** Contrast thresholds are plotted for day 1 (blocks 1 and 2) and day 2 (blocks 3 and 4) of training. Thresholds were calculated using the average contrast of the final 4 reversals per staircase, and averaged across the three interleaved staircases within each block. Coloured lines are individual subject data. Horizontal dashed line represents when the context contrast equalled the target contrast, as in Experiments 2-5 (i.e., signal-to-noise, or target-to-context ratio of one). **B)** Proportion correct plotted per each context contrast level presented, separately for days 1 and 2 (white and grey symbols, respectively). Proportion correct at each contrast level for each day was calculated for each subject, and the average of all subjects is plotted here; the size of the circles represents how many subjects completed trials at a given contrast. Curves were fit using the locally weighted regression (LOESS) model implemented in *ggplot2*. Vertical dashed line represents when the context contrast equalled the target contrast (i.e., the same as the horizontal dashed line in *A*). The horizontal lines represent the average accuracy for textures with a texture-to-context contrast ratio of one, separately for days 1 and 2 (solid and dashed, respectively). These reference points suggest that the magnitude of learning is similar in the current experiment compared to Experiments 3-5. **C)** The change in response accuracy from pre- to post-training for the 6 different conditions. Bars represents averages, and individual symbols represent individual subjects. Error bars in A & C represent ±1 SEM, and shaded regions in B represent 95% confidence intervals of the LOESS fitted line.

manipulation that was used during training in Experiment 6 produced very similar learning as when training with constant contrast and difficult target+context textures: training with target+context textures produces context generalizable learning that is specific to the trained targets. This result is important because the staircase paradigm was more effective in promoting learning across all observers while using the same number of trials.

5.8.3 Discussion

Experiment 6 was designed to test the efficacy of a novel training paradigm, where participants learn to tolerate increasing amounts of non-informative context with practice. The results confirm that this training paradigm was effective at producing significant learning, and the learning showed the same characteristics as previous experiments: learning to identify horizontal targets embedded in a vertical context generalizes to the same targets presented alone, but not to novel horizontal or vertical targets, regardless of context presence.

The results are consistent with a theory in which the decision variable relies on re-weighted information from orientation-selective channels in the visual system. Learning was observed for the trained horizontal targets, with no evidence of transfer to novel horizontal targets, suggesting that the re-weighting of orientation-selective channels is specific to the channels involved for the trained targets, and not a general re-weighting of all channels tuned to horizontally oriented structures.

Comparison of the effects of training in Experiment 6 to the effects obtained in Experiments 3-5 revealed only one significant difference: the change in response accuracy from pre- to post-training for the novel vertical targets embedded within a horizontal context was smaller in Experiment 6 than in the previous experiments. In fact, comparison of the $vT_N + hC$ data in Figures 5.6C to Figures 5.4B & 5.5B reveals that response accuracy slightly decreased after training in Experiment 6, compared to the negligible or slight increases in previous experiments. If the decrease is reproducible, it attests to the effectiveness of the staircase training paradigm in promoting better extraction of horizontal structure: If observers are learning to increase the weight of horizontally tuned channels, then relying on the same channels when the texture's horizontal structure is the non-diagnostic context, as is the case for $vT_N + hC$ stimuli, then performance should decline after training. Experiment 6 is the first out of four experiments where the data trend towards a performance decrement, but it is the

experiment with the most effective training paradigm. We will continue to explore this prediction of the re-weighting hypothesis in future studies.

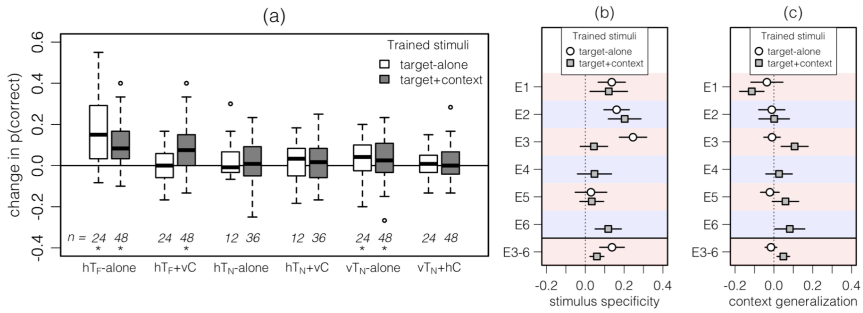

Figure 5.7 – **A)** Boxplots summarizing the change from pre- to post-training in all testing conditions, collapsed across Experiments 3-6. White boxes refer to target-alone trained groups, and grey bars refer to target+context trained groups. The number of participants in each testing condition is indicated under each box, and an asterisk reflects $p < 0.05$ from a one-tailed t test on the pre-post difference. **B)** Stimulus-specificity (x-axis) plotted for each experiment (y-axis) and training group, and also collapsed for Experiments 3-6 (bottom), which were all of the experiments in which the context varied from trial to trial. Error bars represent one-tailed 95% confidence intervals, as we were only interested in whether the magnitude of stimulus-specificity was greater than zero. Stimulus-specificity was statistically significant if the 95% confidence interval did not include 0 (dashed line). **C)** Same as B but for context-generalization.

5.9 Summary of Experiments 3-6

Changes in pre- to post-training response accuracy for all testing conditions are combined across Experiments 3-6 and presented in Figure 5.7A. The pooled data highlight two key results. First, training to identify target-alone textures produced improvements for those six targets presented alone, but learning did not generalize to those six targets embedded within a context, or to novel horizontal targets regardless of the presence of the context. Second, training to identify target+context textures produced improvements to the trained targets embedded within a context and, critically, to the trained targets presented alone. This generalization of learning to the familiar targets presented in a new (i.e., absent) context occurred only in the target+context group, suggesting that identifying targets embedded in a context included some selective learning of the horizontal target structure independent of the context. We also plot stimulus-specificity and context-generalization for all experiments in Figure 5.7B and

5.7C, respectively. Stimulus-specificity was defined for the target-alone group as the difference in the effect of training between hT_F-alone and vT_N-alone, and for the target+context trained group as the difference in the effect of training between $hT_F + vC$ and $vT_N + hC$ textures. Similarly, context-generalization was defined for the target-alone group as the difference in the effect of training between $hT_F + vC$ and $vT_N + hC$, and for the target+context trained group as the difference in the effect of training between hT_F-alone and vT_N-alone textures. The most critical finding here is that when collapsed across the experiments in which the context was variable (i.e., Experiments 3-6), context-generalization was significant for the target+context trained group but not for the target-alone trained group.

5.10 General Discussion

In a series of experiments, we trained participants to identify a set of horizontally-filtered textures in one of two conditions: one in which the texture was presented by itself (target-alone) and one in which it was presented in a non-informative, vertically-filtered context (target+context). Before and after training, we assessed response accuracy to familiar and novel targets presented alone or with a context. In Experiment 1, we found that it was more difficult to identify target+context stimuli than target-alone stimuli, but that they both produced stimulus-specific learning that did not generalize to the untrained context. In Experiment 2, we replicated Experiment 1 while ruling out possible effects of target contrast differences between the texture types. In Experiment 3, we changed the non-informative context in target+context textures across all items on every trial. We found that response accuracy dropped drastically, but learning generalized to the trained targets in the absence of the context. That is to say, individuals trained on target+context textures recognized the familiar targets when presented alone (i.e., "context-generalization"). We replicated this finding in Experiment 4, which also suggested that context-generalization did not apply to novel targets with the same orientation as the trained targets. We confirmed the conclusions of Experiment 4 in Experiment 5, and also ruled out the role of difficulty in promoting context-generalization. That is, training on very difficult target-alone textures did not produce the context-generalization that training with target+context textures did, suggesting the role of the context in producing context-generalization was not confounded with the difficulty of the task. Lastly, Experiment 6 confirmed the results

of Experiment 3-5 using a training paradigm that produced learning in all subjects. In summary, our results indicate that learning to identify horizontal targets embedded within a varying, non-informative vertical context can lead to a strategy of discovering familiar structures in novel contexts or when presented alone, but that the learning is tied to the specific structures that were seen during training.

In some ways, the impact of a non-informative but highly visible context on task performance was surprising. In Experiments 3-5, the target+context group exhibited relatively small amounts of learning even though we excluded participants who failed to show any improvement during training. However, given the consistent effects across Experiments 3-6, and the use of a training paradigm that reliably promoted learning in the participants of Experiment 6, we believe that the small effects of stimulus-specificity and context-generalization are real effects. These results suggest that we can push human perception to differentially rely on oriented structure in a stimulus in which observers must discover which orientations do and do not convey task-relevant information. Follow-up experiments should aim to study how extended training will affect the pattern of results, and if there are more subtle cues we can provide individuals to promote faster learning.

Why did the presence of a non-informative context affect the identification of an orthogonally oriented target structure? One interpretation of the results is that the participants perform the texture identification task using a decision variable that receives input from all orientation channels. Initially, the decision variable equally relies on all orientation channels. This is supported by the relative difficulty of the three different texture types used during the training phase. The easiest textures were target-alone textures which contained diagnostic horizontal structure and no visible structure in the vertical orientations. Hence, vertical channels provided no input to the decision variable, and therefore there was no vertical signal to affect the perceptual decision. In the case of target+context textures with a variable texture (Experiments 3-6), on every trial, the non-diagnostic vertical structure was visible and novel, and thus the decision variable would partly rely on a non-diagnostic input by placing equal weight on horizontally tuned and vertically tuned channels. Consistent with this idea, we see that response accuracy was much worse for target+context textures than target-alone textures. In-between these two texture types is the target+context textures with a constant context (Experiments 1-2). In these instances, equal weighting of all orientations should not be detrimental to performance because, although the vertical

structure is non-informative, it is constant for all textures. Since the non-informative vertical structure is constant from the stimulus to all six alternatives on the response screen, then the stimulus shares some structure with all possible alternatives; hence, in a noisy system such as our visual system, it is conceivable that these textures may lead to matches to the incorrect response alternatives given that all five incorrect alternative textures share *some* signal with the stimulus texture. Furthermore, if the decision variable relies on a signal in which all orientation channels are pooled, then each target will produce a unique percept that remains constant across trials, and thus distinct from all other combinations of the same context with different targets. Therefore, the task can be done with relative ease by relying on a pooled signal.

In addition to differences in overall response accuracy, the different types of textures lead to differences in target-specific and context-specific learning: target-alone and target+(constant context) textures produced target- and context-specific learning, whereas target+(variable context) textures produced target-specific but context-generalizable learning. What produces such a pattern of results? Well, on every trial, in conditions in which the context varies randomly, the horizontal target structure is one of six possible alternatives, whereas the vertical context structure is novel on every presentation. Our results suggest that, with training, the decision variable re-weights its inputs from orientation channels to rely more heavily on information carried by (the reliable and less variable) horizontally tuned channels than (the unreliable and highly variable) non-informative vertically tuned channels. The low variability of the target allowed for the observer to discover idiosyncrasies of the horizontally oriented structures, whereas the high variability of the context prevented any learning of specific idiosyncrasies of vertically oriented structures. Therefore, re-weighting of channels carrying horizontal structure gives rise to context-generalization, wherein the trained decision variable would be able to recognize familiar horizontally oriented structures regardless of the status of the vertically oriented channels. Given that the results from Experiments 4-6 indicate improvements are target-specific, then re-weighting is unlikely to reflect a general increased weight of horizontal channels, but rather, a subset of the channels that carry the signal of the learned targets (e.g., a specific set of *spatially-localized*, horizontally-tuned channels). This suggests that our results are similar to previous results in which observers learned to discriminate using the most diagnostic spatial regions (Gold et al., 2004). Furthermore, a re-weighting account of our results reflects increasing weights of horizontal channels only, and not

decreased weight of vertical channels. If vertical channels were down-weighted, then we would expect a reduction in identification accuracy to vertical target-alone textures. Instead, in all experiments, we do not see any significant decreases in response accuracy to vertical target-alone textures. This is also consistent with a perceptual learning mechanism in which improvements are due to a more efficient signal extraction than a reduction in internal noise (Gold et al., 1999).

If channel re-weighting only acts on the target-relevant channels, then why did training on target-alone textures not produce context generalizable learning? The reverse-hierarchy model of perceptual learning (Hochstein and Ahissar, 2002) posits that perception automatically uses a high-order signal, and with need, recedes to using lower level signals. In the current paradigm, the presence of a context instills the need to use the visual signal at a lower level, where the input is represented across separate orientation channels. Without the context, perceptual learning occurs on a higher order signal where orientation information is not independent. Hence, although the channels carrying the contextual signal do not undergo re-weighting, they are required for the re-weighting of the target-carrying channels. Critically, for the context to effectively push perception to a lower level signal, the context needs to be adequately disruptive, supported by the fact that we did not see evidence for context-generalization when the context did not vary on every presentation. The use of variability in the context is an effective way to prevent the observer from learning specific structure of the context – that is, a basic Hebbian learning machine would become more sensitive to the target structure but not the context structure by mere repetition of the former and no repetition of the latter. In some related work, Hussain et al. demonstrated that when variability is incorporated into the entire texture (i.e., the entire texture is target structure but is novel on every trial), observers are slow at learning to discriminate this class of textures, but the learning is versatile: that is, the *task* is to identify novel targets on every trial, and by default, when the task is successfully learned, then the skill is transferred to novel textures that belong to the same class of stimuli (Hussain et al., 2012a).

McGovern et al. (2012) provided an alternative to the reverse hierarchy model: perception is indeed hierarchical, but learning will transfer to similar tasks in either direction along the hierarchy. For instance, they demonstrate that training on one of three tasks (global form, curvature, or local orientation discrimination) produces transfer to the other two tasks, and the amount of transfer was predicted by the relative

complexity of the transfer task to the training task. That is, training on the low complexity (orientation) or high complex (global form) task both produced the largest transfer to the medium complexity (curvature) task, and less-so to the high or low complexity tasks, respectively. Interestingly, contrary to the authors' prediction that the medium complexity task would transfer to both the low and high complexity tasks, their own data show that the medium complexity task transferred only to the low complexity but not the high complexity task. Nonetheless, McGovern et al. suggest that transfer between two tasks is determined by their proximity to each other along the hierarchy, in either direction. Importantly, the tasks must be along the same hierarchy: for instance, perceiving local orientation elements is part of perceiving a curvature, and perceiving curvature is part perceiving the global form of a shape. In our paradigm, perceiving target-alone and target+context textures should, then, produce transfer to each other, assuming they are close enough along the hierarchy. Regardless of their proximity, the fact that we get asymmetrical transfer suggests the complexity model (McGovern et al., 2012) of perceptual learning is inadequate for our results.

Our results are not consistent with a precision account of transfer (Jeter et al., 2009), in which transfer is expected whenever the participant switches from a high precision (HP) to a low precision (LP) task, but no transfer is expected when switching from a LP to HP task. Texture identification of target-alone stimuli here is likely a HP task, given that there is little uncertainty about the diagnostic structure of the six target-alone stimuli. Identification of the target+context textures is likely a LP task, since the appearance of the diagnostic target structure is made ambiguous by the presence of the non-informative context. Thus, the precision account would predict that training on the HP target-alone texture identification should generalize to the LP target+context texture identification task, and that training on the LP target+context texture identification should not generalize to the HP target-alone texture identification task. We did not see either of these results, and our results were in fact in the opposite direction. Nonetheless, the precision account (Jeter et al., 2009), to our knowledge, has only been tested on simple orientation discrimination tasks, and our predictions to complex identification tasks are only speculations.

Our results also are connected to the study of face perception, in which it has been established that face identity information is significantly more prevalent in the horizontal orientation than vertical orientation band (Dakin and Watt, 2009; Pachai

et al., 2013), and critically, individuals vary significantly in how efficiently they extract the relevant horizontal structure in a face identification task (Pachai et al., 2013). Faces are a special class of stimuli in which humans have had a lifelong opportunity to learn the characteristics of the stimulus class, and therefore, it makes sense that we develop, with experience, a more efficient strategy to process the information from the rich stimulus more effectively (Goffaux et al., 2015). Interestingly, even though healthy adults have already developed a "horizontal bias" in face processing that applies to novel faces (Goffaux et al., 2015; Dakin and Watt, 2009; Pachai et al., 2013), they have an even stronger horizontal bias for personally-familiar faces (Pachai et al., 2017), suggesting that through perceptual learning, the visual system is grows more efficient at processing the diagnostic information of repeated stimuli. The study of orientation biases using non-face stimuli such as textures can inform us on how such perceptual biases develop with experience under fine psychophysical control. In addition, situations where discovering initially ambiguous diagnostic structure in a field of non-diagnostic structure is relevant to gaining expertise in diagnostic imaging. For example, studies in dermatology (Norman et al., 1989) and mammography (Nodine et al., 1999) indicate a significant role of perception in becoming an expert, and therefore simulating such scenarios in the lab using textures may aid in developing effective training paradigms.

Conclusion

Individuals were trained to discriminate textures that only contained horizontally oriented target structure, or horizontally oriented target structure and vertically oriented contextual structure. When the context remained constant, learning was highly specific to the trained context. When the context was novel on every trial, learning generalized to the trained targets without the context, suggesting observers were learning, selectively, the horizontally oriented target embedded in a texture that contained information in all orientations. We did not see convincing evidence that the learning transferred to novel targets in the trained orientation, suggesting that context-generalization does not reflect a break from stimulus specific perceptual learning, but rather represents an instance where stimulus specific perceptual learning becomes robust against changes in the context. We suggest that observers can learn to make perceptual decisions by relying more on certain orientations of information.

This implies that, unlike previous reports, information from independent orientation channels can be represented at the decision level.

References

Ahissar, M. and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–384.

Ball, K. and Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6).

Bennett, P., Hashemi, A., and Sekuler, A. (2015). Position-specific learning in a texture identification task. *Journal of Vision*, 15(12):1135.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–6.

Censor, N., Karni, A., and Sagi, D. (2006). A link between perceptual learning, adaptation and sleep. *Vision Research*, 46(23):4071–4.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Dakin, S. and Watt, R. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

DeValois, R. L. and DeValois, K. K. (1990). *Spatial vision*, volume 31. Oxford University Press.

Diamond, R. and Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2):107–17.

Dosher, B. A. and Lu, Z.-L. L. (1999). Mechanisms of perceptual learning. *Vision Research*, 39:3197–3221.

Fahle, M., Edelman, S., and Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, 35(21):3003–3013.

Fiorentini, A. and Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Research*, 21.

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

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

Gauthier, I. and Tarr, M. J. (1997). Becoming a 'Greeble' expert: Exploring mechanisms for face recognition. *Vision Research*, 37(12):1673–1682.

Goffaux, V., Poncin, A., and Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 Year Old). *PLoS ONE*, 10(9):1–17.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, 402:176–8.

Gold, J. M., Sekuler, A. B., and Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28:167–207.

Hochstein, S. and Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804.

Hussain, Z., Bennett, P. J., and Sekuler, A. B. (2012a). Versatile perceptual learning of textures after variable exposures. *Vision Research*, 61:89–94.

Hussain, Z., McGraw, P. V., Sekuler, A. B., and Bennett, P. J. (2012b). The rapid emergence of stimulus specific perceptual learning. *Frontiers in Psychology*, 3(July):226.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009a). Contrast-reversal abolishes perceptual learning. *Journal of Vision*, 9:1–8.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009b). How much practice is needed to produce perceptual learning? *Vision Research*, 49(21):2624–34.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009c). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, 49(18):2273–84.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2011). Superior Identification of Familiar Visual Patterns a Year After Learning. *Psychological Science*, 22(May):724–30.

Jeter, P. E., Dosher, B. A., Liu, S.-H., and Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research*, 50(19):1928–40.

Jeter, P. E., Dosher, B. A., and Lu, Z.-l. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision*, 9(3):1–13.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(NOV):1–12.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2):467–477.

Mangini, M. C. and Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, 28(2):209–226.

McGovern, D., Webb, B., and Peirce, J. (2012). Transfer of perceptual learning between different visual tasks. *Journal of Vision*, 12:1–11.

McKee, S. P. and Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception & Psychophysics*, 24(3):258–262.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., and Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, 6(10):575–585.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., and Muzzin, L. J. (1989). The Development of Expertise in Dermatology. *Archives of Dermatology*, 125(8):1063–1068.

Olejnik, S. and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, 8(4):434–447.

Olzak, L. A. and Thomas, J. P. (1999). Neural recoding in human pattern vision: Model and mechanisms. *Vision Research*, 39(2):231–256.

Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2018a). The effect of training with inverted faces on the selective use of horizontal structure. *Vision Research*.

Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2018b). The bandwidth of diagnostic horizontal structure for face identification. *Perception*, January 19:1–17.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to Information Conveyed by Horizontal Contours is Correlated with Face Identification Accuracy. *Frontiers in Psychology*, 4(74).

Pachai, M. V., Sekuler, A. B., Bennett, P. J., Schyns, P. G., and Ramon, M. (2017). Personal familiarity enhances sensitivity to horizontal structure during processing of face identity. *Journal of Vision*, 17(6):5.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics Transforming numbers into movies. *Spatial Vision*.

Peterson, M. and Eckstein, M. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109(48).

Petrov, A. A., Dosher, B. A., and Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, 112(4):715–743.

Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256(5059):1018–21.

R Core Team and R Development Core Team, R. (2017). R: A Language and Environment for Statistical Computing.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., and Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, 13(3):250–257.

Rousselet, G. A., Ince, R. A. A., van Rijsbergen, N. J., and Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *Journal of Vision*, 14(13):1–24.

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research*, 51(13):1552–66.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–6.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1):141–145.

Chapter 6

Extended training of orientation filtered textures increases generalization of learning

6.1 Abstract

Textures can be manipulated to contain diagnostic structure (target) in one orientation band, and non-diagnostic structure (context) in a perpendicular orientation band. Identifying targets is easier when targets are presented alone than when presented within contexts, suggesting that the orthogonal but irrelevant information cannot be ignored during perceptual decisions. In previous experiments, we found that learning to identify targets did not transfer to the same targets presented within contexts, but learning to identify targets within contexts did transfer to the same targets when presented alone. However, although learning that occurred with targets embedded in contexts generalized to a new context, the learning did not transfer to novel targets. One limitation of our previous studies is that the magnitude of learning was very small, and therefore we may not have had adequate power to measure transfer to novel stimuli. Here, we addressed this issue by significantly increasing the amount of training. Twelve participants received 4600 or 4800 practice trials in a 1-of-6 texture identification task. The stimuli were horizontally-filtered textures embedded in a non-informative vertical context. Before and after training, we assessed identification accuracy on trained and novel targets with and without contexts. We also tested accuracy on textures where

the target and context orientations were swapped, using both novel and trained targets. The amount of improvement during training varied widely across observers. After training, response accuracy improved significantly for familiar targets presented with and without context, for familiar targets without context but rotated 90 deg, and for *novel* horizontal targets presented alone or within contexts. The transfer of learning was not related in any simple way to the learning rate during training. Our results indicate that perceptual learning of orientation filtered textures varies significantly across individuals, but that it can be generalized to novel and familiar targets in novel contexts. We discuss these results in models of perceptual learning, as well as their implications for perceptual learning in applied settings in which generalization of learning is a critical component of training.

6.2 Introduction

Performance in most perceptual tasks improves with practice, but performance improvements often are restricted to the trained stimuli (Sagi, 2011). Stimulus-specificity is seen as a hallmark of perceptual learning, and is found in diverse tasks including vernier acuity (Fahle and Edelman, 1993), the discrimination of grating orientation and spatial frequency (Fiorentini and Berardi, 1981), the discrimination of motion direction (Ball and Sekuler, 1987), and the identification of textures (Hussain et al., 2009) and faces (Hussain et al., 2012).

Recently, we studied perceptual learning in a texture identification task that used stimuli that contained diagnostic structure in a relatively narrow range of orientations (Chapter 5). Specifically, the textures contained horizontal diagnostic structure that was presented alone or embedded in non-informative, vertical structure (i.e., context). These stimuli (see Figure 6.1) allowed us to ask if observers could learn to extract diagnostic, oriented structure for accurate identification, and, if so, to determine the specificity of the resulting learning. Our working hypothesis was that forcing subjects to discover informative structure would result in learning that generalized more to novel stimuli. Participants practiced in a 1-of-6 identification task with targets either presented alone or within a context. The task was significantly more difficult when targets were presented with a context than when presented alone. Importantly, if observers practiced with targets within a context, they showed complete transfer of learning to the same targets when presented alone (i.e., context generalization).

The opposite was not true: practice with targets presented alone did not produce improvements to the same targets presented within a context. The asymmetric transfer of learning points to a possibly important distinction in learning processes between training with targets presented alone or targets presented within a context. Still, although target+context training led to context-generalization, the generalization did not apply to novel horizontal targets. Failure to transfer learning to novel horizontal targets was against our working hypothesis, so we explore this failure of transfer in this study.

Results from Chapter 5 are limited by the fact that identification performance was still rather low after 960 trials of training. There are two problems with minimal learning in an already difficult task: (1) partial transfer to novel targets could have occurred but fallen below detectable levels, and (2) participants did not have adequate opportunity to actually 'discover' how to best identify these structure. In the current study, we aimed to address these two issues by providing participants with significantly more practice.

We asked participants to practice identifying horizontal targets within a vertical context for up to 4800 trials. Before and after training, we tested identification accuracy in several conditions that allowed us to assess generalization of learning to novel contexts, targets, and orientations. To anticipate our results, we found that participants varied widely in how much they learned, but all showed the same pattern of context generalization that was found in our previous experiments. However, unlike previous results, extended training produced learning that transferred to novel targets. Finally, we found that learning rate was not associated with context generalization, or with generalization to novel targets.

6.3 Methods

6.3.1 Subjects

Twelve adults (4 male; range = 17-27 years old, M = 20.6, SD = 2.97) from McMaster University participated in the experiment. All participants had normal or corrected-to-normal Snellen acuity, and an average Pelli-Robson contrast sensitivity of 1.91 (range = 1.80-1.95, SD = 0.068). All participants were naïve with regard to the experimental hypotheses and had not participated in any of our previous experiments.

6.3.2 Apparatus & Stimuli

The apparatus was the same as the one described in Chapter 5. Stimulus generation was the same as well, but is described below to include some additional conditions.

Textures were generated by first applying an isotropic, bandpass (2-4 cycles per image) ideal spatial frequency filter to 18 256×256 pixel (4.6×4.6 deg of visual angle) patches of Gaussian white noise. Next, 60 deg bandwidth ideal orientation filters centered on either 0 deg (horizontal) or 90 deg (vertical) were used to generate target textures for different conditions. Twelve textures were filtered to preserve horizontally-oriented structure (hT-alone). Six of these textures were were shown during the training and test phases of the experiment and therefore were designated as familiar stimuli (hT_F-alone), and the remaining six textures were used only in the test phase of the experiment to test for transfer to novel targets of the same orientation (hT_N-alone). The six familiar targets were also rotated 90 deg to test for orientation specificity of learning for familiar targets (vT_F-alone). The remaining original six textures were filtered to preserve vertical structure, and they were used as control stimuli: novel targets in the untrained orientation (vT_N-alone). The four sets of target-alone textures were then used to create target+context stimuli. The $hT_{F/N}$-alone textures were combined with vertically filtered (60 deg bandwidth) structure from a random noise patch that was first filtered with an isotropic spatial frequency filter described above. This structure is referred to as the context. The combination of hT-alone and a vertical context formed the $hT_F + vC$ and $hT_N + vC$ stimuli. Similarly, $vT_{F/N}$-alone textures were combined with a horizontally oriented context, forming the $vT_F + hC$ and $vT_N + hC$ textures. Critically, on every trial, and for every response screen alternative, the context was randomly generated and thus contained no task-relevant information. In summary, target-alone textures contained structure *either* in a 60 deg horizontally- or vertically-orientated band of orientations, contained no structure at all other orientations, and all visible structure was diagnostic for the identification task. In contrast, target+context textures contained horizontal *and* vertical structure, but only one band of orientations contained diagnostic structure belonging to the target (i.e., horizontal in $hT_{F/N} + vC$ stimuli and vertical in $vT_{F/N} + hC$ stimuli) whereas the other band of orientations contained non-diagnostic structure referred to as the context.

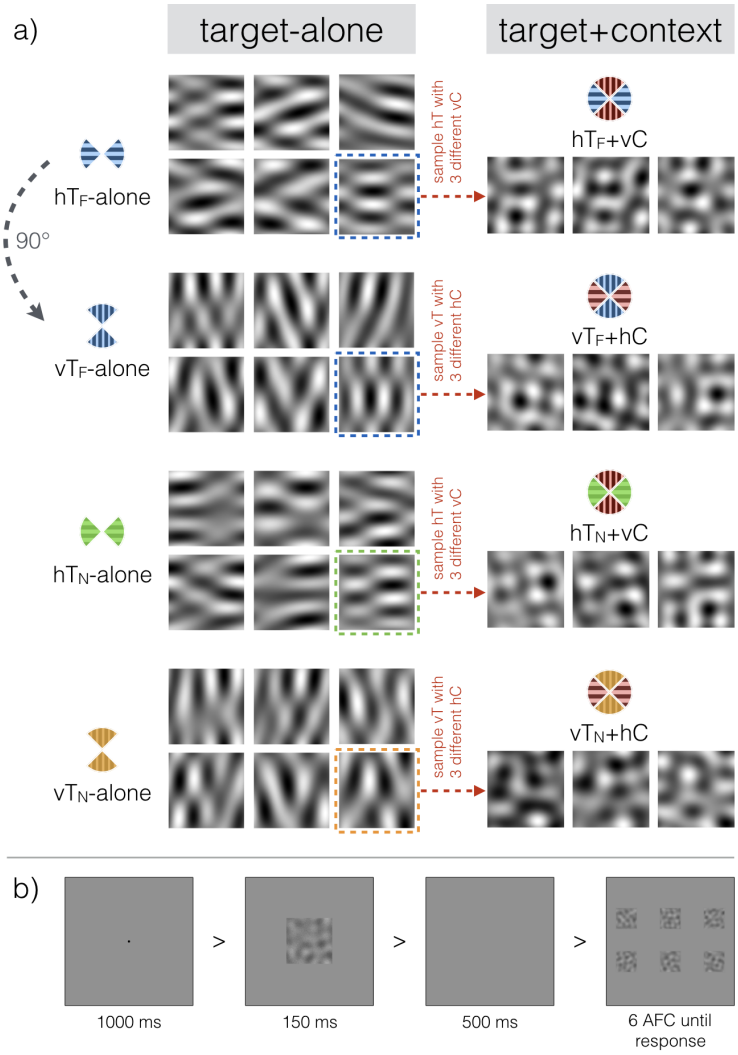

Figure 6.1 – **A)** Orientation filtered textures used in the experiment. Target-alone textures (left) were used as they appear here. Target+context textures (right) are presented only as a demonstration: since there is a virtually infinite number of possible target and context combinations, we demonstrate what a single target appeared like when combined with three of the infinitely possible contexts. Despite the three target+context stimuli all containing identical target structures, the final textures look remarkably different; the difference in appearance is due only to the different contexts. In the actual experiment, a new context was generated on every trial, which was itself unique from the six novel contexts generated on every trial for the response screen. In other words, the stimulus and the correct response matched only in the target and differed in context. **B)** The trial schematic used during both testing and training trials. During training trials, the stimuli were only from the $hT_F + vC$ set.

6.3.3 Procedure & Design

Participants completed a 1-of-6 texture identification task. Every trial followed the same sequence described in detail in Chapter 5 and illustrated in Figure 6.1. During training trials, the stimulus was always selected randomly from the $hT_F + vC$ texture set. The spatial arrangement of the six targets on the response screen remained constant across all training trials. The test phase consisted of two blocks of trials: In one block, the type of stimulus on each trial was selected randomly from four types of target-alone textures (hT_F-alone, hT_N-alone, vT_F-alone, and vT_N-alone), and in the other block the stimulus was selected randomly from four types of target+context textures ($hT_F + vC$, $hT_N + vC$, $vT_F + hC$, and $vT_N + hC$). Within each stimulus set, the spatial arrangement of stimuli on the response screen remained constant across trials belonging to each texture set. Because different texture sets were intermixed within a block of test trials, the items on the response screen changed across test trials.

Each training session consisted of 600 trials with $hT_F + vC$ textures. Each of the six unique targets was repeated 100 times, with every repetition using a new, randomly-generated context. Based on their availability, participants completed 7 or 8 training sessions, for a total of 4200 (n=6) or 4800 (n=6) trials. Testing sessions were completed before and after training. Each testing session consisted of 360 trials (6 texture sets $\times$ 6 targets per texture set $\times$ 10 repetitions per target) Hence, pre- and post-training response accuracy was calculated on 60 trials per condition per testing session (Figure 6.2). Two participants (LUC and SAR) did not complete the vT_F-alone and vT_F+hC test conditions because these conditions were added to the experiment after LUC and SAR had begun training.

After each training session, the experimenter addressed two questions to each participant. The questions were: "How did you find the task today?" and "What were you looking for when identifying textures?". The interviews were informal and were aimed to get a sense of how participants perceived the textures, and if they were able to accurately describe the horizontal structures that were consistent across trials. We took note of when a participant described their strategy as looking for horizontal and diagonal structure. Often times, participants described the horizontal/diagonal structure by gesturing the shape of the targets. We were careful to not ask leading questions, mention horizontal/vertical/diagonal, indicate what they should look for, or acknowledge if their strategy was correct or incorrect. At some point during training,

all but two participants (SOT and KAT) described the stimuli in terms of oriented – i.e, horizontal – structure.

6.4 Results

Statistical analyses were performed with R (R Core Team, 2018), and figures were plotted using the *ggplot2* package (Wickham, 2009).

Response accuracy from individual participants is plotted against training trials in Figure 6.2A. Response accuracy improved for all participants, but the magnitude of learning varied drastically between participants. For each participant, linear regression was used to estimate the best-fitting line to the training data, and the learning rate (i.e., slope) and intercept were recorded. Learning rates ($M = 0.009$, $t_{11} = 5.45$, $p < 0.001$, $d = 2.23$) and intercepts ($M = 0.287$, $t_{11} = 14.8$, $p < 0.001$, $d = 6.05$) were significantly different from zero. Higher intercepts were associated with lower learning rates, but the correlation was not significant ($r = -0.314$, $t_{10} = -1.04$, $p = 0.320$).

Response accuracy during testing trials is plotted in Figure 6.2B. Single subject performance is plotted in colour (top three rows), and group means are plotted in black (bottom row). Within each plot, a positive slope from pre- to post-training reflects an improvement in response accuracy. Learning rates during training differed dramatically across participants (Figure 6.2A), but inspection of Figure 6.2B suggests that individual differences were less pronounced in the test sessions. The effect of training on performance in the test conditions was examined by comparing pre- to post-training response accuracy in each condition using paired sample t tests. The p values were adjusted to maintain a false discovery rate less-than-or-equal to 5% (Benjamini and Hochberg, 1995). Response accuracy improved significantly for familiar horizontal targets presented with vertical contexts ($hT_F + vC$: $M = 0.264$, $t_{11} = 8.86$, $p_{adj} < 0.001$, $d = 2.56$) and with the context removed (hT_F-alone: $M = 0.160$, $t_{11} = 5.47$, $p_{adj} < 0.001$, $d = 1.58$). That is, participants who were trained to identify $hT_F + vC$ textures performed better with those same patterns, but showed context generalization by also performing better with hT_F-alone textures. Response accuracy also improved significantly for horizontally-oriented *novel* targets presented alone (hT_N-alone: $M = 0.146$, $t_{11} = 3.84$, $p_{adj} = 0.007$, $d = 1.11$) and with vertical contexts ($hT_N + vC$: $M = 0.100$, $t_{11} = 2.85$, $p_{adj} = 0.025$, $d = 0.82$), reflecting transfer to targets that they had not previously been trained on. This result, which was not

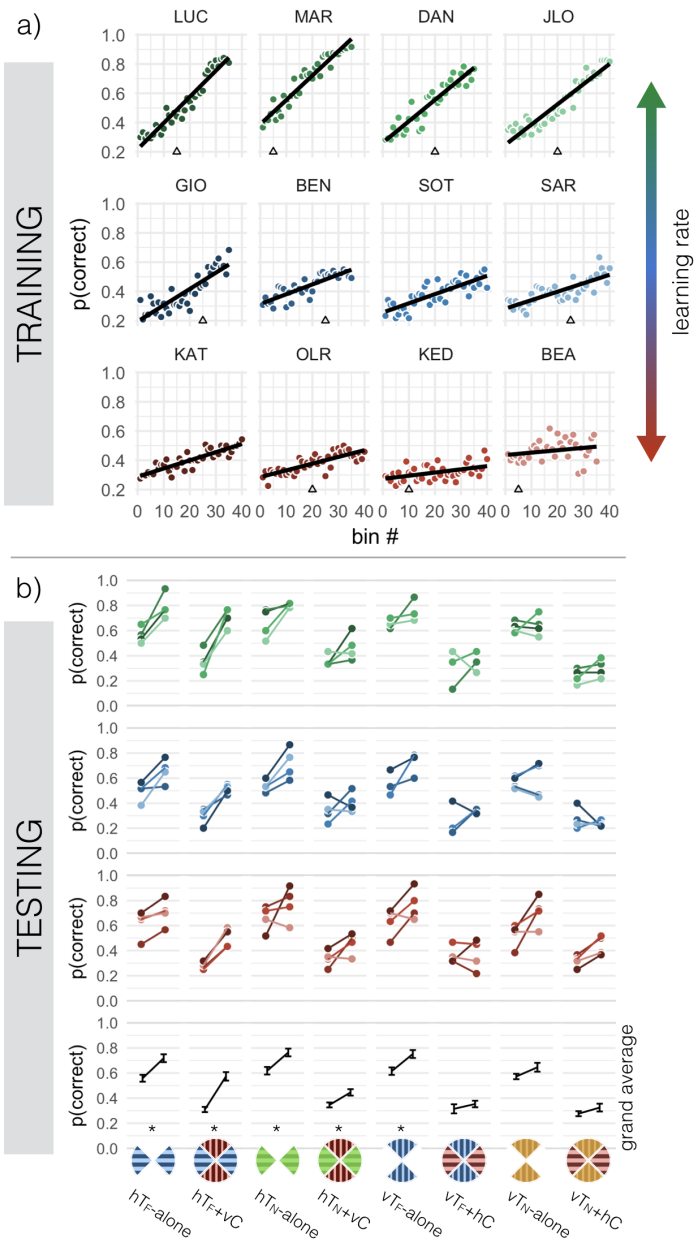

Figure 6.2 – **A)** Training results: single subject response accuracy across training trials, calculated in bins of 120 trials each. Every 5 consecutive bins compose a single session. Response accuracy is plotted separately and arranged in order from the highest to the lowest learning rate. A triangle in each plot (except SOT and KAT) indicates after which bin a participant was able to approximately describe the horizontal targets in post-session interviews. **B)** Testing results: single subject response accuracy at pre- and post-training for each testing condition indicated at the bottom. Four single subjects are grouped into each plot, according to their learning rate (highest learning rate in top row). The fourth row contains the mean across all subjects (±1 SEM). The top three rows suggest that there were no clear differences between generalization of learning across different learning rates. On a group level (bottom row), asterisks correspond to a significant improvements in response accuracy from pre- to post-training.

obtained in Chapter 5, is the first time that we have seen learning generalize to new, horizontally-oriented targets. Response accuracy also improved significantly for familiar targets rotated by 90 deg when they were presented alone (vT_F-alone: $M = 0.137$, $t_9 = 3.68$, $p_{adj} = 0.010$, $d = 1.16$) but not when they were presented within horizontal contexts ($vT_F + hC$: $M = 0.038$, $t_9 = 0.876$, $p_{adj} = 0.404$, $d = 0.28$), suggesting that learning generalized to a novel orientation, but only when the orientation of the target structure (as opposed to the contextual structure) was unambiguous. Lastly, response accuracy did not improve significantly for vertically-oriented novel targets presented alone (vT_N-alone: $M = 0.074$, $t_{11} = 1.83$, $p_{adj} = 0.125$, $d = 0.53$) or within horizontal contexts ($vT_N + hC$: $M = 0.049$, $t_{11} = 1.66$, $p_{adj} = 0.143$, $d = 0.48$). These last two comparisons indicate that the improvements found in other conditions reflect the effects of learning that are tied to characteristics of the stimuli, and are not due to general improvements in performing the task.

To examine the relation between performance during training and testing, we correlated the learning rate and intercept from the learning curves during training with the change in response accuracy in each test condition (Figure 6.3). Learning rates were associated with larger improvements in response accuracy for familiar horizontal targets presented alone (hT_F-alone: $r = 0.650$, $t_{10} = 2.71$, $p_{adj} = 0.165$) and within vertical contexts ($hT_F + vC$: $r = 0.595$, $t_{10} = 2.34$, $p_{adj} = 0.165$), although the correlations were not significant. Learning rate was not significantly correlated with any other testing condition ($-0.27 \leq r \leq 0.16$, $-0.97 \leq t \leq 0.51$, $0.80 \leq p_{adj} \leq 0.94$). The correlation between the the intercept, which represents accuracy at the beginning of training, and improvements at test was largest for novel horizontal targets presented alone (hT_N-alone, $r = -0.554$, $t_{10} = -2.10$, $p_{adj} = 0.493$), although the relationship was not statistically significant. The intercept was not significantly correlated with any other testing condition ($-0.21 \leq r \leq 0.33$, $-0.67 \leq t \leq 0.98$, $p_{adj} = 0.68$). Overall, our measures of performance during the training phase were not strongly associated with improvements in the test conditions.

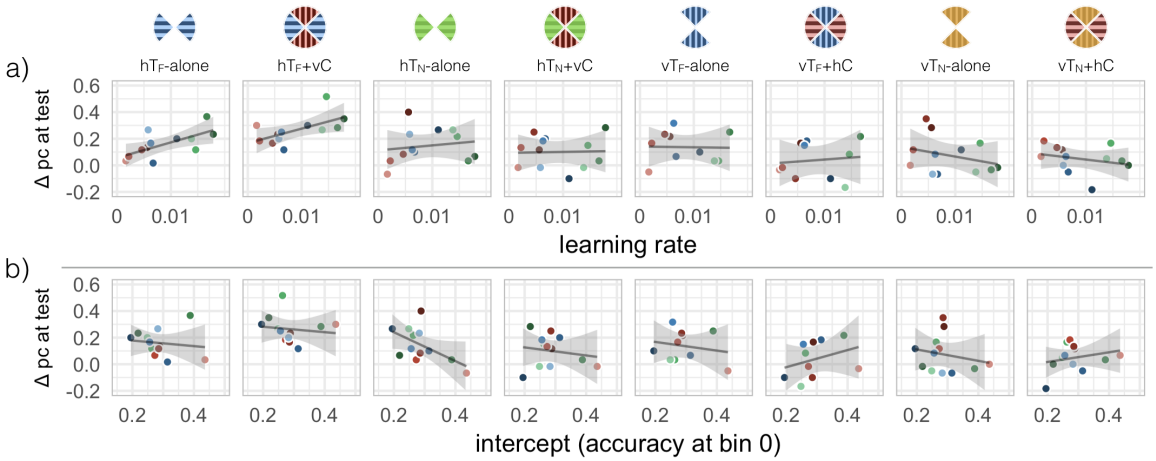

Figure 6.3 – The amount of improvement from pre- to post-training (y-axes) in the various testing conditions is plotted against the (A) slope or (B) intercept of the learning curve during training. Colours correspond to individual subjects from Figure 6.2A. Best fitting lines were estimated using Pearson's least squares, and shaded regions represent the 95% confidence intervals of the correlation. The learning rate (i.e., slope) during training was significantly correlated with improvements in identifying familiar horizontal targets presented alone (hT_F-alone) or with a vertical context (hT_F+vC). The intercept was not correlated with improvement in any testing conditions.

6.5 Discussion

Participants were trained to identify six horizontally-oriented targets embedded within vertically-oriented contexts for 4600 or 4800 trials. Changes in response accuracy during training varied widely across participants, but the variation was not associated with the difference between pre- and post-training accuracy in the test conditions. In fact, after training, all participants showed evidence of context generalization – improved performance for targets presented *without* context – and most participants showed stimulus generalization – improved performance for novel targets in the trained orientation and familiar targets in a novel orientation. Our results indicate that learning to identify textures can be transferrable to variations of the trained targets, and to novel targets sharing some key characteristics with the trained target. We will discuss these key characteristics below.

The finding of context generalization is consistent with previous findings from experiments that used many fewer training trials (Chapter 5). One interpretation is that perceptual learning of textures is due to a decision variable re-weighing its incoming connections (Petrov et al., 2005) from different orientation channels (De

Valois et al., 1982) to improve the signal-to-noise ratio (Gold et al., 1999). Such a mechanism explains why training to identify a set of horizontal targets embedded in vertical contexts produces learning that transfers to horizontal targets presented alone. Throughout training, channels tuned to horizontally-oriented inputs transmit signals belonging to a set limited to six targets, whereas channels tuned to vertically oriented inputs transmit signals belonging to an infinite set of contexts, and thus never repeated. A basic Hebbian learning rule would predict that the decision variable would strengthen its connections from the horizontally-tuned but not vertically-tuned inputs. Interestingly, as the experiments in Chapter 5 demonstrate, training to identify horizontal targets in the absence of vertical contexts *does not* transfer to the same targets embedded in vertical contexts. Therefore, additional constraints, such as those in the reverse hierarchy model of learning (Hochstein and Ahissar, 2002) are needed to understand the pattern of transfer thus far. Better identification of target+context textures requires the decision variable to rely differently on horizontally- and vertically-oriented signals, whereas better identification of target-alone textures does not require the decision variable to distinguish the signal according to orientation, since non-target orientations contain no structure and thus are not disruptive to the integrity of the target's representation.

Extensive training with horizontal targets embedded in vertical contexts produced learning that transferred to novel horizontal targets presented alone or within vertical contexts. This is the first time that we have been able to show significant transfer to novel targets. The re-weighting hypothesis used to account for context generalization may also apply here. Target-specific context generalization, as that seen in Chapter 5, may be due to the re-weighting of orientation channels that are spatially-selective, therefore learning will not generalize to similarly oriented structures that fall on different spatial receptive fields. Non-specific context-generalization, i.e., transfer to novel but horizontally oriented targets, may reflect the re-weighting of inputs from a level of orientation-selective channels that are not spatially selective (i.e., the signal is pooled across space but not orientation). Alternatively, instead of generalization explained by the pooling of spatial information, it is possible that extended training led to the re-weighting of many more channels, thus covering a larger spatial receptive field. We cannot distinguish these two possibilities using our results.

Evidence of transfer following extensive training but not following shorter training regimens as in Chapter 5 can reflect one of, or a combination of, two possibilities. First,

transfer effects may grow slowly but steadily with increasing training, and therefore they simply may have been too small to measure reliably in previous experiments that used fewer training trials. Alternatively, transfer effects may emerge rapidly and suddenly, following a criterion number of training trials. In this view, we failed to observe transfer in previous experiments because we did not provide participants with an adequate opportunity to discover techniques/strategies that transfer to novel stimuli. In the current study, the experimenter did an informal follow-up after each training session to learn how participants described the manner in which they performed the task. Indeed, 10 of the 12 participants described an "aha! moment", which was when they first said that they were looking for "horizontal lines" or something similar. The first session following a participant's report of horizontal structure is indicated by a triangle in Figure 6.2A. Response accuracy in these sessions ranged between 32% and 56% ($M = 44.9\%$, $SD = 7.6\%$), possibly indicating that a minimum accuracy was required for participants to discover the underlying target structure. However, inspection of Figure 6.2 suggests that the timing of the realization that the targets consisted of horizontal structure did not coincide with any meaningful changes in learning rate. Also, the two participants (SOT & KAT) who never said that they were searching for horizontal structure nevertheless exhibited context and stimulus generalization during the test phase that was similar to that found in the other subjects. Although obtaining and analyzing "aha! moments" was informal and not conclusive, it does suggest that becoming aware of, and being able to verbalize, the orientation of the diagnostic structure in our stimuli was not related in any obvious way to the process(es) underlying improvements in this task. A more formal and detailed investigation into how participants describe the targets and how well they perform is needed to better understand the relationship between psychological experiences and perceptual abilities.

In addition to context- and target-generalization, participants showed transfer of learning to familiar targets that were rotated by 90 degrees. That is, participants recognized vertical targets presented alone (vT_F-alone) as being the rotated version of the same horizontal targets they were trained on while they were presented in vertical contexts ($hT_F + vC$). There was no significant transfer to the rotated familiar targets if they were presented within horizontal contexts ($vT_F + hC$), indicating that participants were not able to recognize familiar structures in untrained orientation bands if the orientation of the targets are not explicitly visible in the stimulus. In other words, participants learned that when a texture contains information in both

horizontal and vertical orientation bands, they should put emphasis on the processing of horizontal structure. This is not a good strategy when the otherwise familiar target is now vertically oriented unbeknownst to the participant. Transfer to familiar but rotated targets may be an extension of work suggesting that the bandwidth of orientation channels are adjustable (Taylor et al., 2014); perhaps, in addition to the adjustment of the bandwidth, the orientation itself can be adjusted following adequate perceptual learning. Learning may reflect orientation-specific re-weighting of inputs to the decision variable, but the decision variable appropriately relies on previously untrained orientations if and when the target structure's orientation is visible and unambiguous to the observer. Nonetheless, learning appears to be orientation-selective relative to the stimulus, not relative to the observer.

In a separate but related line of research, studies of face perception have highlighted the uneven distribution of information across different orientation bands, with the majority of identity information conveyed by horizontally oriented structures (Pachai et al., 2013). Indeed, humans develop and/or learn to preferentially process horizontal relative to vertical structures (Goffaux et al., 2015; Balas et al., 2015, 2017), and the strength of this preference is predictive of overall face identification abilities (Pachai et al., 2013; Chapter 2). One reason the study of orientation selectivity in faces is informative here is the methodological difference in how orientation selectivity is measured. For example, horizontal facial structure can be presented alone (Dakin and Watt, 2009; Chapter 2) like our target-alone conditions, or within an uninformative facial context (Pachai et al., 2018; Chapter 4) as in our target+context conditions. Other methods include using orientation filtered noise masks (Pachai et al., 2013), or orientation-specific phase scrambling (Jacques et al., 2014). Indeed, these methods have all provided converging evidence that face perception relies heavily on horizontally oriented structures; however, the manipulation in which filtered faces are presented within the context of a non-diagnostic face may be the most valid form of testing the mechanisms underlying face perception, since there are no added cues as to what orientations contain the diagnostic structures. For example, Goffaux and Greenwood (2016) used filtered face alone stimuli to conclude that human face perception relies on a 25 degree bandwidth of horizontally oriented information, whereas Pachai et al. (2018) used filtered faces in an uninformative but facial context to conclude that the bandwidth humans use is much larger at 75 degrees, suggesting that perhaps the unambiguity of what orientation bands contain information in the filtered face alone

conditions may be a cue for information processing; this cue does not typically exist in a face, nor does it exist in filtered faces presented in an uninformative context. Similarly, we found different results for target-alone and target+context textures for familiar targets that were rotated at test. Specifically, we observed transfer to rotated targets when presented alone, but not when presented within a context. This distinction suggests that target+context textures are a valid measure of the observers learned mechanism/strategy for that class of stimuli, whereas target-alone textures are a measure of how sensitive the perceptual system is to stimulus features.

One criticism of our results is that for a supposed learning experiment, improvements in the training task are rather slow. However, in a practical application, medical practitioners such as radiologists and dermatologists have to regularly make perceptual judgements on visual stimuli that may share some characteristics with our task and results. In one study on the development of expertise in dermatologists, Norman et al. (1989) demonstrate that it takes years of experience for medical students, and subsequently practitioners, to increase their diagnosis rate from 21% to 87% accurate (see Nodine et al., 1999, for similar results in mammography). The rate of learning in medical practitioners is significantly improved with feedback (Sowden et al., 1996), indicating that mere visual exposure to images is inadequate for optimal learning. Indeed, this is consistent with the idea that medical practitioners have to discover the relevant characteristics of the stimuli for accurate diagnoses. One interpretation of the difference between our current results and those in Chapter 5 is that the extended training allowed for the possibility that observers could discover the characteristics of the diagnostic structure embedded in the noise. Furthermore, as we suggest that our work is consistent with the reverse-hierarchy model of perceptual learning (Hochstein and Ahissar, 2002), medical expertise acquisition has also been framed in the reverse hierarchy model (Taylor, 2007). That is, with enough practice with difficult stimuli, perception of to-be medical imaging experts begins to rely on low-level rather than high-level representations of the stimulus, allowing for increased sensitivity to subtle but diagnostic features that are not well represented in high-level representations. Linking our controlled, in lab results with that of medical imaging experts not only increases the importance of our research, but also indicates that we may be able to use these constructed images to mimic, and possibly enhance, the training regimen of medical imaging students.

In summary, our results indicate that extensive practice in identifying orientation-

manipulated textures produces transferrable learning. The learning transfers to textures that are similar to the trained targets in at least one of two ways: novel targets that share the same orientations as trained targets, or novel orientations that have the same but rotated structure as the trained targets. By studying perceptual learning using complex textures with the diagnostic value of orientation structure manipulated, we can better understand the dynamics of perceptual learning in every day life that goes beyond simple stimuli such as gabors, verniers, or random dot kinematograms. If similar perceptual learning underlies expert human face perception, then understanding its dynamics can be informative to promoting efficient and expert perception in general.

References

Balas, B. J., Schmidt, J., and Saville, A. (2015). A face detection bias for horizontal orientations develops in middle childhood. *Frontiers in Psychology*, 06(June):1–8.

Balas, B. J., van Lamsweerde, A. E., Saville, A., and Schmidt, J. (2017). School-age children's neural sensitivity to horizontal orientation energy in faces. *Developmental Psychobiology*, 59(7):899–909.

Ball, K. and Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision research*, 27(6).

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57(1):289–300.

Dakin, S. C. and Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

De Valois, R. L., William Yund, E., and Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22(5):531–544.

Fahle, M. and Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, 33(3):397–412.

Fiorentini, A. and Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision research*, 21.

Goffaux, V. and Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, 6:34204.

Goffaux, V., Poncin, A., and Schiltz, C. (2015). Selectivity of face perception to horizontal information over lifespan (from 6 to 74 Year Old). *PLoS ONE*, 10(9):1–17.

Gold, J. M., Bennett, P. J., and Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, 402:176–8.

Hochstein, S. and Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804.

Hussain, Z., McGraw, P. V., Sekuler, A. B., and Bennett, P. J. (2012). The rapid emergence of stimulus specific perceptual learning. *Frontiers in psychology*, 3(July):226.

Hussain, Z., Sekuler, A. B., and Bennett, P. J. (2009). Contrast-reversal abolishes perceptual learning. *Journal of Vision*, 9:1–8.

Jacques, C., Schiltz, C., and Goffaux, V. (2014). Face perception is tuned to horizontal orientation in the N170 time window. *Journal of vision*, 14:1–18.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., and Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, 6(10):575–585.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., and Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125(8):1063–1068.

Pachai, M. V., Bennett, P. J., and Sekuler, A. B. (2018). The bandwidth of diagnostic horizontal structure for face identification. *Perception*, January 19:1–17.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to information conveyed by horizontal contours is correlated with face identification accuracy. *Frontiers in psychology*, 4(74).

Petrov, A. A., Dosher, B. A., and Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, 112(4):715–743.

R Core Team (2018). R: A Language and Environment for Statistical Computing.

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision research*, 51(13):1552–66.

Sowden, P. T., Pauli, R., Roling, P., Davies, I. R. L., and Hammond, S. M. (1996). Improvement in screening performance: the importance of appropriate feedback in screening mammography. *Proc. SPIE*, 2712(March 1996):180–188.

Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2014). Evidence for adjustable bandwidth orientation channels. *Frontiers in Psychology*, 5(JUN):1–10.

Taylor, P. M. (2007). A review of research into the development of radiologic expertise: Implications for computer-based training. *Academic Radiology*, 14(10):1252–1263.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Chapter 7

General Discussion

Perceptual learning occurs because observers become more sensitive to informative aspects of the stimuli (Sagi, 2011). For example, response accuracy in a texture identification task improves because observers learn to base their decisions on diagnostic regions (Gold et al., 2004). Lifelong experience with faces also is associated with observers basing their decisions on diagnostic regions (Sekuler et al., 2004). In texture identification, the diagnostic regions differ between textures, and therefore different observers rely on different regions. On the other hand, faces are relatively uniform in their spatial arrangement, so most observers rely on the eye and brow regions for face identification. Despite these differences, the informative aspects of the identity of textures and faces are accessible and apparent to observers.

In other situations the informative aspects of an image may not be apparent. For instance, medical imaging experts often search for subtle cues in an otherwise ambiguous and uninformative stimulus (e.g., Norman et al., 1989; Nodine et al., 1999). Perceptual learning certainly has a role in the development of expertise, but expertise is not simply explained by improvements in sensitivity to the low contrast features of an image (Sowden et al., 2000). One possibility is that experts have *discovered* what diagnostic information looks like, and the process of discovery has allowed them to better identify the diagnostic aspects of new images of the same type.

The work presented in this dissertation shows that learning to identify textures for which the informative aspects must first be discovered may be difficult, but the resulting learning can be generalized to new textures. In the current experiments, the informative and uninformative stimulus structure occurred in different bands of

orientations. The idea of separating informative/non-informative structure in different orientation bands was inspired by research in face perception, wherein observers have a bias to rely on horizontal structure more than vertical structure. Therefore, our initial experiments established that orientation-selective face processing has behavioural and neural markers, and that these markers are associated with identification accuracy. We also showed that age-related deficits in face identification can be reduced with perpetual learning, and the change is associated with the enhancement of horizontal bias.

7.1 Summary of Results

The results in Chapter 2 establish that a behavioural bias for the horizontal structure of a face is associated with a similar bias in the ERP response to faces. Specifically, we demonstrate that several measures of face processing – including identification accuracy, N170 amplitude, and N250 amplitude – were similarly affected by the removal of horizontal, but not vertical, structure. Relative to full faces, faces with no horizontal structure were associated with decreased response accuracy, a decreased N170 amplitude, and a decreased N250 amplitude. Full face identification accuracy was correlated with behavioural and N250 horizontal bias, but not N170 horizontal bias, suggesting that processes underlying the N250 may be critical for face identification.

The results in Chapter 3 replicate the findings of Chapter 2, but with older adults. Similar to what was seen in younger adults, older adults exhibited behavioural, N170, and N250 sensitivity to differently oriented structure in faces. Critically, horizontal biases were decreased in older relative to younger adults, but were still significantly correlated with face identification accuracy. Unlike what was found with younger adults, face identification accuracy in older adults was significantly correlated with the N170 horizontal bias. This result was interpreted as evidence that early visual processes limit face identification more in older adults relative to younger adults.

The results in Chapter 4 show that age-related changes in face identification can be reduced with perceptual learning. Perceptual learning of faces was associated with a significant increase in horizontal bias. Post-training, but not pre-training, horizontal bias was correlated with face identification accuracy, suggesting that older adults face identification is associated with horizontally biased processing of facial structure. The results suggest that perceptual learning can be used to enhance orientation-selective

processing of stimuli for which orientation-selective processing is advantageous.

In Chapter 5, the idea of discovering orientation-selective processing for a novel stimulus class was explored. Stimuli in a texture identification task were modified to contain informative structure in a horizontal orientation band, and non-informative structure in a vertical orientation band. In these textures, the informative structure was not easily apparent to the observer, since the non-informative structure, and therefore the appearance of the entire texture, changed on every trial. The results show that learning to identify these types of textures is incredibly slow, but that learning generalizes to textures that did *not* contain non-informative structure. Observers were able to recognize familiar informative structure when presented without the uninformative structure. The opposite was not true: training to identify informative structure presented alone did not generalize to the same informative structure embedded in uninformative structure. Hence, practice with textures that promoted the discovery of information in certain orientation bands was effective in producing generalization, although the learning did not transfer to novel structures.

Lastly, the results in Chapter 6 extend the findings of Chapter 5 by showing that a considerable increase in the amount of practice improved the generalizability of learning. Indeed, these experiments found, for the first time, that perceptual learning in this task generalized to novel textures. The rate of learning during training varied significantly between observers, but the rate of learning was not related to generalization in any obvious way. The results are consistent with the idea that prolonged training provided observers with more opportunities to discover the relevant structure, and that the discovery of the relevant structure enables transfer to novel textures where the relevant structure can be discovered in a similar way.

7.2 Implications

7.2.1 Face perception

The results presented in this dissertation have implications in several areas of vision science. One major implication is in our understanding of face perception. Previous research established that horizontal structure is necessary for accurate face identification (Dakin and Watt, 2009), and that individual differences in horizontal bias are associated with face identification accuracy (Pachai et al., 2013), and even that horizontal structure

is needed for the behavioural (Goffaux and Dakin, 2010) and N170 (Jacques et al., 2014) face inversion effects. In Chapter 2, we demonstrate that horizontal structure is necessary for a complete, face-like neural response: whereas the neural response was largely unaffected with the removal of vertical structure, the N170 and the N250 were both significantly affected by the removal of horizontal structure. The N170 has been suggested to be associated with eye detection (Rousselet et al., 2014), consistent with our results showing the N170 amplitude decreasing significantly when information about the eyes, which is conveyed by horizontal structure, is removed. According to the Lateral Inhibition, Face Template, and Eye Detector (LIFTED) model of face processing (Nemrodov et al., 2014), the N170 is the response of an eye-detector which is inhibited by the presence of non-eye features in the parafoveal region. That is, the N170 is larger to isolated eyes than to eyes within a face because isolated eyes are not inhibited by surrounding features. Our results do not directly address this theory, because our filtering manipulation is agnostic to any specific feature, and therefore all filtered faces contain at least some information about all facial features. However, the fact that filtering is agnostic to features is one advantage of studying faces using orientation filtering rather than manipulating subjectively defined features of a face (e.g., Piepers and Robbins, 2012). Studying faces using the orientation structure is advantageous because orientation filtering is consistent with the representation of the signal at early visual processing stages (Hubel and Wiesel, 1968; Pantle and Sekuler, 1968; Blakemore and Campbell, 1969; De Valois et al., 1982a,b). More recently, activity in the fusiform face area has also been shown to be driven by horizontally oriented facial structure (Goffaux et al., 2016).

The results also help understand the characteristics of age-related changes in face identification (Grady et al., 1994). In Chapter 3, we show that a significant amount of age-related changes in face identification are accounted for by changes in the horizontal biases of earlier and later visual processes reflected in the N170 and N250, respectively, as well as changes in the behavioural horizontal bias not accounted for by the two neural markers. However, we show that after accounting for all three measures of horizontal bias, there was still a significant effect of age on face identification accuracy. This result suggests that changes in horizontal bias do not entirely account for age-related changes in face identification accuracy, so other factors need to be considered.

Finally, older adults are only one of many populations who show deficits in face perception. In Chapter 4, we show that practice with upright faces increases the

magnitude of horizontal bias in older adults. Other populations with deficits in face perception include individuals with prosopagnosia (Bodamer, 1947; Hecaen and Angelergues, 1962; Barton, 2003) and autism (Langdell, 1981; Davies et al., 1994; Rutherford et al., 2007). Horizontal bias is shown to be attenuated in individuals with prosopagnosia (Pachai et al., 2015), but horizontal bias of individuals with autism has not yet been measured. It is unclear if a similar training paradigm as that used with older adults will aide face perception in individuals with prosopagnosia or autism, but training regimens focused on affected horizontal bias may prove to be effective. It is unclear if practice-induced changes in horizontal bias measured with one set of faces will transfer to another set of faces. Training regimens can consider the results from Chapters 5 and 6 and incorporate appropriate variability in non-informative stimulus aspects to help promote generalizable learning.

7.2.2 Perceptual learning

The results presented in this dissertation contribute to our understanding of the mechanisms of perceptual learning. The results in Chapter 5 provide support for the reverse hierarchy model of perception (Ahissar and Hochstein, 2004), in which perception relies on the discriminable signal at the highest processing level along the visual pathway. In summary, training to identify targets embedded in an uninformative context generalized to identifying familiar targets presented without a context, whereas training to identify targets presented without a context did not generalize to identifying familiar targets embedded in an uninformative context. Critically, for the asymmetric generalization to occur, the uninformative context needed to be different in every exemplar. This finding suggests that for perception to rely on a signal lower in the processing hierarchy, observers must not be given an opportunity to learn idiosyncrasies of the uninformative context. Alternatively, the observers discover that the context is uninformative based on its variability, thereby allowing them to ignore the context. In any case, learning to base decisions on information from certain orientation bands indicates that past failures to to get observers to base decisions on individual orientations (e.g., Olzak and Thomas, 1991; Olzak and Wickens, 1997) may be overcome with an appropriate perceptual learning paradigm.

An important implication of our results concerns the role of discovery in the generalization of perceptual learning. Instead of perceptual learning occurring by improving sensitivity to informative aspects that are specific to the set of trained

stimuli, our results suggest that perceptual learning may (also) be a consequence of learning to discover task-relevant structure. Learning to discover structure based on some general properties (e.g., horizontal instead of vertical structure) may be a powerful way to use limited sets of stimuli in a perceptual task during training and have learning generalize to untrained stimuli. Interestingly, whereas stimulus-specificity of perceptual learning has been shown to increase with increased training durations (Jeter et al., 2010; Hashemi et al., 2013), the results in Chapter 6 suggest that substantially longer training durations allow for more opportunities to discover the relevant structure. Hence, the characteristics of perceptual learning may differ when the the task requires observers to learn how to discover informative structure rather than the specifics of a a small stimulus set. This proposal is consistent with the idea that the stimulus-specificity of learning is not defined by *where* along the visual pathway the locus of learning is, but rather *what* is learned (Mollon and Danilova, 1996). Our results from Chapters 5 and 6 suggest that the *what* that is learned is which channels in the visual system are most likely to convey the task-relevant structure.

Better identification of stimuli with task-relevant structure hidden within task-irrelevant structure is akin to the development of expertise in the interpretation of medical images. Although Chapter 6 showed that you need thousands of trials to reach relatively high accuracy levels, under 5000 trials in two weeks of training is relatively little compared to the years of experience required by doctors to become diagnostic imaging experts (Norman et al., 1989; Nodine et al., 1999). Using stimuli similar to the textures used in this dissertation, or creating novel stimuli using image statistics of different types of medical images, may be a good first step in developing relatively brief training paradigms for medical students. Furthermore, the individual differences in learning during the texture identification training task, if coupled with a link between performance in a texture identification task and the ability to identify anomalies in medical images, may aide in distinguishing prospective medical imaging experts based on their performance on a standardized test administered early in their career.

7.3 Future research

The experiments presented in this dissertation provide results motivating several lines of future research. Chapter 1 shows that face-related neural responses, namely the N170 and N250, are sensitive to horizontal facial structure, but that only the N250 horizontal

bias correlates with human face identification accuracy. Follow-up experiments should assess if the N250 horizontal bias is modulated by task-relevancy; given that the N250 is sensitive to face identity, does it show a horizontal bias even when face identity is irrelevant to the task? Does the N250 horizontal bias correlate with performance on other tasks, such as emotion, gender, or viewpoint discrimination? Duncan et al. (2017) recently showed that horizontal structure is informative in an emotion discrimination task for most emotions, and also that horizontal bias was associated with emotion discrimination. How horizontal bias is reflected in the ERP to non-identification tasks is still unclear.

The results in Chapters 2 and 3 may underestimate the effect of stimulus information on the behavioural and neural measures of horizontal bias. Future studies should instead filter stimuli like Chapter 4 so that observers are unaware of what orientation bands convey task-relevant information, and therefore will employ a processing strategy that is not influenced by the filtering condition. The prediction is that such stimuli should reduce, or abolish, the horizontal bias at the N170, and therefore allow us to examine the horizontal bias of the N250 without residual effects from the N170. This prediction is justified by results showing that the N170 is sensitive to face, or specifically eye, information (e.g., Rousselet et al., 2014), but not to face identity (Amihai et al., 2011). The stimuli in Chapter 4 always resemble a full face with intact eyes, and therefore the informativeness of the eye in the different filter conditions is no longer confounded with the presence of eye structure.

Chapter 4 provides positive results for practice-based increases in face identification accuracy and horizontal bias in older adults. For these results to be applicable as a clinical tool, practice-based improvements must generalize to novel faces, and therefore follow-up studies should evaluate the stimulus-specificity of the results presented in Chapter 4. Furthermore, future studies should aim to use stimuli that are more inclusive and reflective of the sample's population. For instance, our experiment requires older adults to identify younger adult faces, which may not be reflective of the faces they see in their every day life. Indeed, face recognition is better for own-age than other-age faces (Wiese et al., 2013), and for own-race than other-race faces (Malpass and Kravitz, 1969). Together with the fact that recognition and horizontal bias are better for familiar than unfamiliar faces (Pachai et al., 2017), age-related deficits in horizontal bias need to be evaluated using faces belonging to older adults.

The results in Chapters 5 and 6 are the first demonstrations that the perceptual

learning of textures may be accomplished by learning how to discover task-relevant information. Similar to Experiment 6 in Chapter 5, future studies should focus on identifying training paradigms that reliably produce learning. For instance, one way to potentially make the task easier is by reducing the variability in the contextual structure. Whereas Experiments 1 and 2 in Chapter 5 used the same context on the stimulus and response screen, as well as between trials, and Experiments 3 to 6 used a unique context from stimulus to response screen, as well as between trials, future studies should test the effect of using unique contexts between trials, but the same context within each trial (stimulus and response screen). In such cases, the task should become easier since the correct response will better resemble the stimulus, but observers will not be given the opportunity to repeatedly see the context, and therefore learning will not be tied to context structure (as was seen in Experiments 1 and 2).

Orientation-selective processing of textures, as demonstrated in Chapters 5 and 6, share basic similarities with the horizontal bias seen in face perception. However, several differences exist, and future studies can address these differences. One major difference is that the information in natural images of faces is not exclusive to specific orientation bands, despite being more informative in horizontal relative to vertical orientation bands. One way to address this is to create textures using jittered filter bandwidths such that on any given trial, informative structure is conveyed by a horizontally-centred orientation band with a ± 10 deg to ± 90 deg bandwidth. The orthogonal bandwidth would then contain uninformative structure, so that observers are unaware of the filtering condition and therefore need to discover the task-relevant structure with practice. With these stimuli, the optimal strategy would be to rely most on structure conveyed by a ± 10 deg orientation band, since the structure will always be informative, and less and less on structure conveyed by orientations beyond this bandwidth. This type of graded horizontal bias is akin to the horizontal bias seen in face perception (e.g., Chapters 2-4).

Identifying similarities in orientation-selective processing between faces and textures will also enable us to better understand the neural correlates of face perception and orientation-selective processing. The neural response to faces has been well-studied, including evidence that face-like neural responses to non-face objects can be achieved through training (Rossion et al., 2002). Therefore, achieving expertise with the textures used in this dissertation may then allow us to test the orientation-selectivity of specific neural responses of interest (N170 and N250) to non-face images of expertise. Such

studies could begin to distinguish if horizontal bias in face processing is, at a neural level, similar to a learned horizontal bias to non-face stimuli. Recent results have suggested that horizontal bias to faces originates in the fusiform face area (Goffaux et al., 2016), so unless lab-based expertise with textures recruits the fusiform face area, it is possible that the two are not governed by the same neural mechanism.

7.4 Conclusion

This dissertation provides evidence of orientation-selective processing in the human visual system. The results build on previous research on face perception by establishing a neural sensitivity to information conveyed by different orientation bands, and also that orientation-selective processing of faces is affected by perceptual learning. Additionally, a novel stimulus set is created and used to demonstrate that orientation-selective processing can be learned, and that successful learning may hinge on an observer's ability to learn how to discover task-relevant information when such information is not easily distinguishable. Learning to discover task-relevant information is an effective way to promote a generalizable form of perceptual learning, which is a welcome change from decades of research pointing to an often stimulus-specific form of perceptual learning. In summary, this dissertation demonstrates that under certain circumstances, information conveyed by different orientation bands can be used differentially by the human observer.

References

Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10):457–464.

Amihai, I., Deouell, L. Y., and Bentin, S. (2011). Neural adaptation is related to face repetition irrespective of identity: A reappraisal of the N170 effect. *Experimental Brain Research*, 209(2):193–204.

Barton, J. J. (2003). Disorders of face perception and recognition. *Neurologic Clinics*, 21(2):521–548.

Blakemore, C. and Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, 203(1):237–260.

Bodamer, J. (1947). Die prosop-agnosie. *Archiv für Psychiatrie und Nervenkrankheiten*, 179(1-2):6–53.

Dakin, S. C. and Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, 9:1–10.

Davies, S., Bishop, D., Manstead, A. S. R., and Tantam, D. (1994). Face Perception in Children with Autism and Asperger's Syndrome. *Journal of Child Psychology and Psychiatry*, 35(6):1033–1057.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982a). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5):545–559.

De Valois, R. L., Yund, E. W., and Hepler, N. (1982b). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22(5):531–544.

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., and Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision*, 17(14):7.

Goffaux, V. and Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, 1(September):1–14.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., and Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, 81:1–11.

Gold, J. M., Sekuler, A. B., and Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28:167–207.

Grady, C. L., Ma Maisog, J., Horwitz, B., Ungerleider, L. G., Mentis, M. J., Salerno, J. A., Pietrini, P., Wagner, E., Haxby, J. V., Gillette, J., Giacometti, K., Baldwin, P., Jacobs, G., Stein, S., Green, S., Fluck, S., and Der, M. (1994). Age-related changes in cortical blood flow activation during visual processing of faces and location. *The Journal of Neuroscience*, 14(3):1450–1462.

Hashemi, A., Lass, J. W., Truong, D., Sekuler, A. B., and Bennett, P. J. (2013). The time-course of rapid stimulus-specific perceptual learning. *Journal of Vision*, 13(9):1090.

Hecaen, H. and Angelergues, R. (1962). Agnosia for Faces (Prosopagnosia). *Archives of Neurology*, 7(2):92–100.

Hubel, D. and Wiesel, T. (1968). Receptive Fields and Functional Architecture of Monkey Striate Cortex. *The Journal of physiology*, 195(1):215–243.

Jacques, C., Schiltz, C., and Goffaux, V. (2014). Face perception is tuned to horizontal orientation in the N170 time window. *Journal of Vision*, 14:1–18.

Jeter, P. E., Dosher, B. A., Liu, S.-H., and Lu, Z.-L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research*, 50(19):1928–40.

Langdell, T. (1981). *Face perception: An approach to the study of autism*. PhD thesis, University of London.

Malpass, R. S. and Kravitz, J. (1969). Recognition for faces of own and other race faces. *Journal of Personality and Social Psychology*, 13(4):330–334.

Mollon, J. D. and Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision*, 10(1):51–58.

Nemrodov, D., Anderson, T., Preston, F., and Itier, R. J. (2014). Early sensitivity for eyes within faces: A new neuronal account of holistic and featural processing. *NeuroImage*, 97:81–94.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., and Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, 6(10):575–585.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., and Muzzin, L. J. (1989). The Development of Expertise in Dermatology. *Archives of Dermatology*, 125(8):1063–1068.

Olzak, L. A. and Thomas, J. P. (1991). When orthogonal orientations are not processed independently. *Vision Research*, 31(1):51–57.

Olzak, L. A. and Wickens, T. D. (1997). Discrimination of complex patterns: Orientation information is integrated across spatial scale; spatial-frequency and contrast information are not. *Perception*, 26(9):1101–1120.

Pachai, M. V., Bennett, P. J., Sekuler, A. B., Corrow, S., and Barton, J. J. S. (2015). Sensitivity to Horizontal Structure and Face Identification in Developmental Prosopagnosia and Healthy Aging. *Canadian Journal of Experimental Psychology*, 69(4):341.

Pachai, M. V., Sekuler, A. B., and Bennett, P. J. (2013). Sensitivity to information conveyed by horizontal contours is correlated with face identification accuracy. *Frontiers in Psychology*, 4(74).

Pachai, M. V., Sekuler, A. B., Bennett, P. J., Schyns, P. G., and Ramon, M. (2017). Personal familiarity enhances sensitivity to horizontal structure during processing of face identity. *Journal of Vision*, 17(6):5.

Pantle, A. and Sekuler, R. (1968). Size-Detecting Mechanisms in Human Vision. *Science*, 162(3858):1146–1148.

Piepers, D. W. and Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, 3(DEC):1–11.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., and Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, 13(3):250–257.

Rousselet, G. A., Ince, R. A. A., van Rijsbergen, N. J., and Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *Journal of Vision*, 14(13):1–24.

Rutherford, M. D., Clements, K. A., and Sekuler, A. B. (2007). Differences in discrimination of eye and mouth displacement in autism spectrum disorders. *Vision Research*, 47(15):2099–2110.

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research*, 51(13):1552–66.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–6.

Sowden, P. T., Davies, I. R. L., and Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):379–390.

Wiese, H., Komes, J., and Schweinberger, S. R. (2013). Ageing faces in ageing minds: A review on the own-age bias in face recognition. *Visual Cognition*, 21(9-10):1337–1363.

Appendix

F Tables for Chapter 5

Analysis of variance tables, as described in Chapter 5, are presented here in full for the reader's benefit. Results are presented for ANOVAs conducted on the testing data. Analyses of training data, and any t tests or other focussed contrasts are not included in this appendix, as they are well described in the text.

Experiment 1

	df_N	df_D	F	p	η^2_G
Intercept	1	20	53.3	0.000	0.445
trainGroup	1	20	3.80	0.065	0.054
stimNovelty	1	20	1.39	0.252	0.021
testContext	1	20	0.00	0.991	0.000
trainGroup:stimNovelty	1	20	0.66	0.425	0.010
trainGroup:testContext	1	20	17.7	0.001	0.197
stimNovelty:testContext	1	20	1.56	0.226	0.009
trainGroup:stimNovelty:testContext	1	20	44.3	0.000	0.212

Table A1 – Chapter 5 Experiment 1. Data from all participants submitted to a 2 (training group) × 2 (testing context) × 2 (stimulus novelty) ANOVA

	df_N	df_D	F	p	η^2_G
Intercept	1	11	44.5	0.000	0.534
stimNovelty	1	11	2.14	0.171	0.051
testContext	1	11	8.17	0.016	0.179
stimNovelty:testContext	1	11	11.6	0.006	0.136

Table A2 – Chapter 5 Experiment 1. Data from the target-alone training group submitted to a 2 (testing context) × 2 (stimulus novelty) ANOVA

	df_N	df_D	F	p	η^2_G
Intercept	1	9	14.1	0.005	0.341
stimNovelty	1	9	0.06	0.808	0.002
testContext	1	9	10.7	0.010	0.227
stimNovelty:testContext	1	9	55.0	0.000	0.313

Table A3 – Chapter 5 Experiment 1. Data from the target+context training group submitted to a 2 (testing context) × 2 (stimulus novelty) ANOVA

Experiment 2

	df_N	df_D	F	p	η^2_G
Intercept	1	21	53.9	0.000	0.530
trainGroup	1	21	3.98	0.006	0.077
stimNovelty	1	21	20.3	0.000	0.159
testContext	1	21	4.14	0.055	0.034
trainGroup:stimNovelty	1	21	0.09	0.768	0.001
trainGroup:testContext	1	21	26.6	0.000	0.186
stimNovelty:testContext	1	21	0.08	0.787	0.001
trainGroup:stimNovelty:testContext	1	21	21.2	0.000	0.157

Table A4 – Chapter 5 Experiment 2. Data from all participants submitted to a 2 (training group) × 2 (testing context) × 2 (stimulus novelty) ANOVA

	df_N	df_D	F	p	η^2_G
Intercept	1	10	34.4	0.000	0.640
stimNovelty	1	10	8.70	0.015	0.139
testContext	1	10	3.92	0.076	0.075
stimNovelty:testContext	1	10	17.5	0.002	0.139

Table A5 – Chapter 5 Experiment 2. Data from the target-alone training group submitted to a 2 (testing context) × 2 (stimulus novelty) ANOVA

	df_N	df_D	F	p	η^2_G
Intercept	1	11	18.6	0.001	0.381
stimNovelty	1	11	11.8	0.806	0.182
testContext	1	11	32.7	0.000	0.314
stimNovelty:testContext	1	11	8.65	0.013	0.178

Table A6 – Chapter 5 Experiment 2. Data from the target+context training group submitted to a 2 (testing context) × 2 (stimulus novelty) ANOVA

Experiment 3

	df_N	df_D	F	p	η^2_G
Intercept	1	22	60.8	0.000	0.438
trainGroup	1	22	1.61	0.218	0.020
stimNovelty	1	22	22.5	0.000	0.241
testContext	1	22	10.4	0.004	0.102
trainGroup:stimNovelty	1	22	1.08	0.311	0.015
trainGroup:testContext	1	22	16.3	0.001	0.151
stimNovelty:testContext	1	22	28.2	0.000	0.176
trainGroup:stimNovelty:testContext	1	22	10.4	0.004	0.073

Table A7 – Chapter 5 Experiment 3. Data from all participants submitted to a 2 (training group) × 2 (testing context) × 2 (stimulus novelty) ANOVA

	df_N	df_D	F	p	η^2_G
Intercept	1	11	26.9	0.000	0.469
stimNovelty	1	11	20.0	0.001	0.283
testContext	1	11	17.0	0.002	0.326
stimNovelty:testContext	1	11	46.4	0.000	0.317

Table A8 – Chapter 5 Experiment 3. Data from the target-alone training group submitted to a 2 (testing context) × 2 (stimulus novelty) ANOVA

	df_N	df_D	F	p	η^2_G
Intercept	1	11	45.1	0.000	0.404
stimNovelty	1	11	5.90	0.033	0.194
testContext	1	11	0.72	0.413	0.009
stimNovelty:testContext	1	11	1.79	0.208	0.039

Table A9 – Chapter 5 Experiment 3. Data from the target+context training group submitted to a 2 (testing context) × 2 (stimulus novelty) ANOVA

Experiment 4

	df_N	df_D	F	p	η^2_G
Intercept	1	22	15.9	0.001	0.201
trainGroup	1	22	2.06	0.165	0.032
stimNovelty	1	22	0.71	0.409	0.005
testContext	1	22	2.99	0.098	0.030
trainGroup:stimNovelty	1	22	1.21	0.284	0.009
trainGroup:testContext	1	22	0.12	0.737	0.001
stimNovelty:testContext	1	22	0.21	0.654	0.002
trainGroup:stimNovelty:testContext	1	22	0.77	0.388	0.009

Table A10 – Chapter 5 Experiment 4. Data from all participants submitted to a 2 (training group) × 2 (testing context) × 2 (stimulus novelty) ANOVA

Experiment 5

	df_N	df_D	F	p or p_{HF}	η^2_G
Intercept	1	22	4.98	0.036	0.061
trainGroup	1	22	0.00	0.982	0.000
stimType	2	44	1.20	0.310	0.014
testContext	1	22	0.91	0.351	0.007
trainGroup:stimType	2	44	1.09	0.346	0.013
trainGroup:testContext	1	22	4.22	0.052	0.032
stimType:testContext	2	44	1.66	0.203	0.020
trainGroup:stimType:testContext	2	44	0.17	0.847	0.002

Table A11 – Chapter 5 Experiment 5. Data from all participants submitted to a 2 (training group) × 2 (testing context) × 3 (stimulus type) ANOVA

	df_N	df_D	F	p or p_{HF}	η^2_G
Intercept	1	11	1.85	0.201	0.051
stimType	2	22	0.07	0.926	0.002
testContext	1	11	3.49	0.089	0.057
stimType:testContext	2	22	1.50	0.244	0.030

Table A12 – Chapter 5 Experiment 5. Data from the target-alone training group submitted to a 2 (testing context) × 3 (stimulus type) ANOVA

	df_N	df_D	F	p or p_{HF}	η^2_G
Intercept	1	11	3.72	0.080	0.075
stimType	2	22	2.72	0.088	0.060
testContext	1	11	0.86	0.374	0.012
stimType:testContext	2	22	0.37	0.676	0.012

Table A13 – Chapter 5 Experiment 5. Data from the target+context training group submitted to a 2 (testing context) × 3 (stimulus type) ANOVA

Experiment 6

	df_N	df_D	F	p or p_{HF}	η^2_G
Intercept	1	11	6.81	0.024	0.125
stimType	2	22	7.07	0.004	0.140
testContext	1	11	1.23	0.291	0.025
stimType:testContext	2	22	1.10	0.351	0.028

Table A14 – Chapter 5 Experiment 6. Data from all participants submitted to a 2 (testing context) × 3 (stimulus type) ANOVA