On early sensory experience in humans and machines

Dissertation

zur Erlangung des Doktorgrades

des Fachbereichs Humanwissenschaften

der Universität Osnabrück

vorgelegt

von

Marin Vogelsang

aus

Tokio, Japan

Osnabrück, 2023

On early sensory experience in humans and machines

Marin Vogelsang

PhD thesis
Institute of Cognitive Science
University of Osnabrück

Supervisors:
Prof. Dr. Gordon Pipa (University of Osnabrück) and
Prof. Dr. Pawan Sinha (Massachusetts Institute of Technology)

November 2023

ABSTRACT

Human perceptual development typically evolves in a temporally structured manner. As newborns, we start out with limited perceptual abilities and acquire greater and greater proficiencies over the subsequent months or years. In the visual modality, for instance, a newborn experiences the world with poor color sensitivity, resolution acuity, and contrast sensitivity initially. Then, over time, greater visual capabilities are being acquired. While these developmental progressions are well-established, their potential functional significance is not yet determined.

In this thesis, I report computational tests of the hypothesis that initial sensory degradations characteristic of normal human development actively support the acquisition of perceptual proficiencies. In a first study, inspired by the low-frequency prenatal auditory experience of a fetus in the intrauterine environment, I studied the consequences of commencing auditory development with initially low-pass-filtered sounds. In a second study, I examined the impact of beginning visual experience with initially color-degraded inputs. In both cases, simulations with deep neural networks revealed that training with a developmentally inspired progression of inputs, evolving from poor to rich, led to superior generalization performance and the emergence of spatially or temporally extended receptive fields. As part of a third study, I conducted computational explorations of the consequences of the joint temporal progression of visual acuity and color sensitivity during early visual development. Specifically, I tested whether this joint progression may help account for the origin of an important organizational principle of the early visual system: the division into the magnocellular and the parvocellular pathway. My computational results provide support for this hypothesis.

In addition, some of the computational studies are complemented by experiments with children who were born blind and treated for their blindness late in life, as part of the joint humanitarian and scientific initiative Project Prakash. In addition to providing important insights into cortical plasticity late in life, some of these results provide additional support in favor of the hypothesis that initial sensory degradations may be adaptive. Together, this work sheds new light on the functional significance of normal developmental trajectories, helps account for some perceptual deficits reported in individuals whose perceptual experience differs from normal development, and inspires new computational training procedures for deep neural networks.

ZUSAMMENFASSUNG

Die menschliche Wahrnehmungsentwicklung verläuft typischerweise zeitlich strukturiert. Als Neugeborene beginnen wir mit limitierter Wahrnehmung und erwerben im Laufe der folgenden Monate oder Jahre immer bessere Fähigkeiten. Visuell zum Beispiel erlebt ein Neugeborenes die Welt zunächst mit geringer Farbempfindlichkeit, Auflösungsschärfe und Kontrastempfindlichkeit. Im Laufe der Zeit werden dann immer bessere visuelle Fähigkeiten erworben. Während diese Entwicklungsverläufe gut bekannt sind, ist ihre mögliche funktionelle Bedeutung noch nicht geklärt.

In dieser Dissertation untersuche ich computergestützt die Hypothese, dass anfängliche sensorische Limitationen, die für die normale menschliche Entwicklung charakteristisch sind, den Erwerb von Wahrnehmungsfähigkeiten aktiv unterstützen. In einer ersten Studie untersuchte ich, inspiriert von der pränatalen Hörerfahrung eines Fötus in der intrauterinen Umgebung mit niedrigen Frequenzen, die Folgen, auditorische Entwicklung mit anfänglich tiefpassgefilterten Geräuschen zu beginnen. In einer zweiten Studie untersuchte ich die Auswirkungen des Beginns von visueller Erfahrung mit anfänglich farblich degradierten Inputs. In beiden Fällen ergaben Simulationen mit tiefen neuronalen Netzen, dass das Training mit einer von der Entwicklung inspirierten Progression der Inputs, die sich von schwach zu stark entwickelt, zu einer besseren Generalisierung und räumlich oder zeitlich erweiterten rezeptiven Feldern führt. Im Rahmen einer dritten Studie erforschte ich die Folgen der gemeinsamen zeitlichen Entwicklung von Sehschärfe und Farbsensitivität während der frühen visuellen Entwicklung computergestützt. Insbesondere habe ich untersucht, ob diese gemeinsame Progression dazu beitragen kann, den Ursprung eines wichtigen Organisationsprinzips des frühen visuellen Systems zu erklären: die Aufteilung in den magnozellulären und den parvozellulären Pfad. Die Ergebnisse meiner Untersuchung unterstützen diese Hypothese.

Darüber hinaus werden manche der simulationsbasierten Studien durch Experimente mit Kindern ergänzt, die von Geburt an blind sind und im Rahmen der gemeinsamen humanitären und wissenschaftlichen Initiative Project Prakash behandelt wurden. Diese Ergebnisse liefern nicht nur wichtige Einblicke in die kortikale Plastizität im späteren Leben, sondern stützen auch die Hypothese, dass anfängliche sensorische Degradationen adaptiv sein können. Insgesamt wirft diese Arbeit ein neues Licht auf die funktionelle Bedeutung normaler Entwicklungsverläufe, hilft bei der Erklärung einiger Wahrnehmungsdefizite, die bei Personen festgestellt wurden, deren Wahrnehmungserfahrung von der normalen Entwicklung abweicht, und gibt Anregungen für neue computergestützte Trainingsverfahren für tiefe neuronale Netze.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my supervisors, co-authors, friends, and family members who have supported me throughout this thesis project.

First and foremost, I wish to thank Gordon Pipa and Pawan Sinha. I have been extremely fortunate to have been working with you. Your guidance, support, and scientific vision is only surpassed by your incredible kindness. I would further like to thank Sidney Diamond, whose support and contributions have been key to the success of many projects. It is not an exaggeration to state that the three of you are the ideal role models for being not only excellent scientists but also wonderful human beings.

My gratitude extends to the many collaborators and co-authors I have had the pleasure of working with, as well as the entire Project Prakash team for allowing me to be part of this remarkable initiative.

Finally, I would like to thank my friends and family for the support that they have always provided, and Lukas, Jannis, and Priti for proofreading parts of this thesis.

CONTENTS

II Research

Part I

INTRODUCTION

LIST OF CONTRIBUTIONS

PAPERS – MAIN CONTRIBUTIONS TO THIS THESIS:

1. **Vogelsang, M.***, Vogelsang, L.*, Diamond, S., & Sinha, P. (2023). "Prenatal auditory experience and its sequelae". **Published** in Developmental Science, 26(1), e13278. https://doi.org/10.1111/desc.13278. (* = equal contribution)

2. **Vogelsang, M.***, Vogelsang, L.*, Gupta, P.*, Gandhi, T., Shah, P., Swami, P., Gilad-Gutnick, S., Ben-Ami, S., Diamond, S., Ganesh, S., & Sinha, P. (Submitted). "Impact of early visual experience on later usage of color cues". **Under peer-review** at Science (initial submission: September 2023). (* = equal contribution)

3. **Vogelsang, M.**, Vogelsang, L., Pipa, G., Diamond, S., & Sinha, P. (Submitted). "On the origin of the parvo- and magnocellular division: potential role of developmental experience". **Submitted manuscript** (initial submission: October 2023).

4. Vogelsang, L.*, **Vogelsang, M.***, Pipa, G., Diamond, S., & Sinha, P. (Submitted). "Butterfly effects in perceptual development: a review of the 'adaptive initial degradation' hypothesis". **Under peer-review** at Developmental Review (initial submission: May 2023). (* = equal contribution)

PAPERS – ADDITIONAL CONTRIBUTIONS:

5. Gupta, P., Shah, P., Gilad-Gutnick, S., **Vogelsang, M.**, Vogelsang, L., Tiwari, K., Gandhi, T., Ganesh, S., & Sinha, P. (2022). "Development of visual memory capacity following early-onset and extended blindness". **Published** in Psychological Science, 33(6), 847-858. https://doi.org/10.1177/09567976211056664.

6. Bi, S., Chawariya, A., Ganesh, S., Gupta, P., Huang, Y., Jazayeri, K., Kumar, R., Ralekar, C., Singh, C., Tiwary, A., Vogelsang, L., **Vogelsang, M.**, Yadav, M., & Sinha, P. (2023). "Scholastic status of congenitally blind children following sight surgery". **Published** in International Journal of Special Education, 37(2), 160-168. https://doi.org/10.52291/ijse.2022.37.49. (alphabetical author ordering, except for P. Sinha)

7. Gupta, P., Shah, P., Gilad-Gutnick, S., **Vogelsang, M.**, Vogelsang, L., & Sinha, P. (Submitted). "The influence of semantics on visual memory capacity in children and adults". **Under revision** at

British Journal of Developmental Psychology (initial submission: December 2022)

8. Jarudi, I. N., Braun, A., **Vogelsang, M.**, Vogelsang, L., Gilad-Gutnick, S., Bosch, X. B., Dixon, III, W. V., & Sinha, P. (2023). "Recognizing distant faces". **Published** in Vision Research, 205, 108184. https://doi.org/10.1016/j.visres.2023.108184.

CONFERENCE ABSTRACTS:

9. **Vogelsang, M.***, Vogelsang, L.*, Diamond, S., & Sinha, P. (2021). "On prenatal auditory experience in humans and its relevance for machine hearing". **Poster presented** at ICLR Workshop "Generalization beyond the training distribution in brains and machines", 2021, Online. (* = equal contribution)

1 INTRODUCTION

1.1 THESIS OVERVIEW

Human perceptual development typically evolves in a temporally structured fashion. As newborns, we start out with limited perceptual abilities and acquire greater and greater proficiencies over the months and years to follow. In the visual domain, for instance, a newborn begins to experience the world with poor color sensitivity, resolution acuity, and contrast sensitivity initially (Dobkins et al., 1997; Dobson and Teller, 1978; Kiorpes, 2016). Over time, greater perceptual capabilities are being acquired. While these developmental progressions are well-established, little is known about their potential significance in setting up sensory processing strategies early in life that might prove to be beneficial later on.

Systematic examinations of the potential benefits of starting perceptual development with initially degraded sensory inputs form the core of this PhD thesis. These examinations heavily build on the use of deep neural networks as computational model systems. In addition, some of the studies are complemented by experiments with a unique population of atypically-developed individuals – children who were born blind and were treated for their blindness late in life. The latter has been enabled by having had the fortune of working with the joint humanitarian and scientific initiative 'Project Prakash' (Mandavilli, 2006; Sinha, 2013), founded by Prof. Sinha at MIT about twenty years ago. The work presented in this thesis, thus, spans the diverse domains of typical development, atypical development, and computational modeling. In order to better contextualize the contributions described here, I will begin by providing high-level introductions to all three scientific fields (Sections 1.2-1.4) and subsequently bring them together in the specific context of probing the role of early sensory experience (Sections 1.5 and 1.6).

In Section 1.2., I will introduce the scientific study of human perceptual development. I will begin by reviewing some of the field's history, which is deeply linked to century-old philosophical debates about the role of experience in development, and move on to describe the most common methodologies used today for measuring perceptual proficiencies in newborns and infants. This will be followed by a summary of what have been experimentally determined to be the developmental trajectories of a set of particularly relevant perceptual proficiencies. These trajectories provide an important foundation for the work described in later parts of this thesis. Specifically, they motivate the examination of the perceptual capabilities of individuals who experienced deviations from normal developmental trajectories and, moreover, form the basis for incorporating aspects of typical development into the training procedures of computational model systems.

In Section 1.3., after having introduced the study of perceptual development, I will motivate, from both a scientific and a societal perspective, working with individuals who experienced deviations from normal development. While there are many different cases of atypical development, considering the results reported in this thesis, this section will be focused specifically on the case of children who were born blind and who received treatment for their blindness late in life. I will conclude by reviewing some of the past findings that have emerged from studies of such individuals and discuss what these findings may teach us about both typical and atypical development.

In Section 1.4, after having motivated studies of human perceptual development, I will introduce and motivate the use of deep neural networks as computational model systems. For better contextualization, I will begin by providing a short overview of the history of artificial neural networks, from the early work in the first half of the twentieth century to modern-day deep learning systems. This will be followed by a brief introduction to deep convolutional neural networks as well as an assessment of the utility of using such networks as models of the human perceptual system.

In Section 1.5, after having introduced the three fields that are central to this thesis, I will bring them together in the context of the 'Adaptive Initial Degradation' (AID) hypothesis. As noted earlier, key to this hypothesis is the idea that the normal developmental trajectory of perceptual functions, transitioning from limited to proficient, may be adaptive and help, rather than hinder, early perceptual organization. More specifically, initially degraded inputs may induce the developing brain to establish processing mechanisms that are able to stably integrate such degraded inputs, instead of allowing for the development of an over-reliance on the availability of fine-grained details. After describing this idea in greater detail and presenting relevant past work, I will explain why testing the significance of normal developmental trajectories of perceptual function relies on studies of atypically-developed individuals whose perceptual experience did not follow such trajectories, along with computational models that can be trained with different trajectories of sensory experience.

Finally, in Section 1.6., I will describe the further structure of this thesis and provide an overview of the specific contributions reported in this thesis. This summary is split into 'key' and 'additional' contributions. The key contributions thereby comprise three projects focused on computational (and, in part, experimental) examinations of the AID hypothesis in the domains of prenatal hearing (Chapter 2), color vision (Chapter 3), and the joint developmental progression of visual acuity and color sensitivity (Chapter 4). The description of these three projects is followed by a review paper that aims to evaluate and reflect on the AID hypothesis more broadly (Chapter 5).

The additional work reported in the main part of this thesis contains computational contributions to a study testing the visual memory capacity of Prakash individuals immediately and longitudinally following sight-restoring surgeries (Chapter 6), as well as contributions to an examination of the educational performance of the Prakash group (7). Additional work provided in the thesis appendix includes computational contributions to work on the role of semantics in visual memory capability in children and adults (Appendix A) as well as to a study of face recognition at a distance (Appendix B).

1.2 NORMAL PERCEPTUAL DEVELOPMENT

Theoretical considerations as well as empirical examinations of the development of perceptual capabilities from birth to adulthood have played an important role in the history of developmental psychology. In this section, I will begin by briefly reviewing some of that field's history, which dates back to century-old philosophical debates about the role of experience in development but has, over time, witnessed a shift from purely logical considerations to empirical examinations. I will then describe two of the most common methodologies used today for examining perceptual proficiencies in newborns and infants, followed by a summary of what these methods have revealed. This review is highly selective. In light of the work I am reporting in this thesis, I will focus primarily on the visual modality and the development of proficiencies such as visual acuity and color sensitivity in particular. This review will be complemented by a summary of the development of the sensitivity to temporal frequencies in the auditory domain, providing the basis for the work reported in Chapter 2. I will conclude this review by linking it back to the general theme that perceptual proficiencies are initially degraded but improve throughout the developmental timeline, which forms the basis for subsequent parts of this thesis.

1.2.1 *A brief history of the role of experience in perceptual development*

In his book 'The Principles of Psychology', now cited over sixty-thousand times, William James famously described the initial experience of a newborn as a "blooming, buzzing confusion" (James, 1890). Accordingly, the discipline of perceptual development should be concerned with how such initial experience is transformed into full perceptual proficiency as we develop. While there is broad agreement on the significance of experimentally examining the temporal progression of perceptual proficiencies, to what extent a newborn's initial experience really is blooming, buzzing, and confusing lies at the core of a remarkably long-lasting debate among philosophers and psychologists.

As argued in Kellman and Arterberry (2007), James' famed statement can be understood as representing what is termed the 'constructivism' view, dating back to the work of 'empiricist' philosophers such as Berkeley (1709) or Locke (1690). According to this view, initial sensory experience is essentially devoid of meaning and needs to undergo a learning process to become meaningful (Kellman and Arterberry, 2007). Berkeley (1709) has provided one of the classic arguments for this perspective. Specifically, he expressed the concern that visual sensations, comprising information such as the brightness or color at different locations of the retina, are not sufficiently informative for a newborn to infer the size or position of the objects they represent. Considering the infinitely many possibilities to map what has been flatly projected onto the retina back to objects in the three-dimensional world that we live in, such inference would be infeasible. He thus argued that visual sensations, in order to become meaningful, would need to be complemented with information from the non-visual senses, such as through touch or movement (Berkeley, 1709). Other researchers have expressed similar notions of learning through association and experience. For instance, Mill (1865) understood an object as being defined by the set of percepts that would result from observing that object under all possible conditions and viewpoints. As noted in Kellman and Arterberry (2007), this experience-based 'constructivist' perspective, which has dominated the field of developmental psychology from early on, has primarily been based on logical considerations and thought experiments.

In contrast to this perspective is the 'ecological' view, as coined by Gibson (Gibson, 1979; Gibson, 1966). Key to this perspective is the idea that certain regularities and constraints have always governed the physical world we inhabit and that incorporating such constraints into our perceptual machinery would provide an evolutionary advantage. Hering (1861) provided one of the earliest empirical contributions to understanding how such regularities could be incorporated into our perceptual system. Specifically, he described an integrative and potentially innate mechanism through which the information picked up by both eyes is compared to each other, serving the extraction of depth information. As pointed out by Kellman and Arterberry (2007), this finding has an important implication when revisiting the argument promoted by Berkeley (1709). Specifically, Berkeley's concern about the ambiguity of visual information, and the need for relating it to non-visual sensations, appears valid when considering only the information available to a single retina. However, one may also interpret the visual machinery as an integrated system comprising two eyes, linked to a moving body, and equipped with appropriate visual mechanisms for stereopsis and other high-level information. In this case, the system is confronted with less ambiguous information and may not necessarily need to be dependent on experience. This renders the question of whether it actually is experience-dependent, an empirical one (Gibson, 1979; Kellman and Arterberry, 2007).

To conclude, whether it is called 'ecological vs. constructivist', 'nativists vs. empiricist', or 'nature vs. nurture', the debate between the two philosophical camps has been active for centuries. As Daw (2014) argued, neither of the two views alone is correct, considering that some perceptual skills appear to be learned while other proficiencies are available at birth. Other researchers have proposed reconciling the two views (e.g., Haber, 1985; Norman, 2002), and Kellman and Arterberry (2007) call to the empirical sciences to provide more comprehensive accounts. As such, it is important to note that what can significantly inform this long-lasting debate are empirical studies with newborns and infants, focused on the role of experience, and the quality of such experience, relative to the presence of innate biases and pre-programmed maturational processes. With this motivation, I will now describe two of the most common experimental methodologies whose invention has enabled such examinations.

1.2.2 *Methodologies for examining visual proficiencies in newborns and infants*

In light of the significance of empirically examining perceptual capabilities, it is essential to consider what methodologies practically allow for such examination in newborns and infants. Let us consider the example of visual acuity – one of the most fundamental and most commonly tested aspects of visual proficiency. Carrying out such examination with an adult participant would be straightforward. For instance, following one of the most common procedures for testing visual acuity, a participant would be seated 20 feet away from a Snellen chart, which is depicting letters of different sizes, and asked to name the letters shown on the chart. Once the participant is unable to do so, the limit of visibility can be extracted and expressed as a Snellen fraction. Normal vision would thereby correspond to a value of 20/20, whereas a value of, for instance, 20/100 would indicate that the tested person's vision at a distance of 20 feet is as good as that of a person with normal vision at a distance of 100 feet (Azzam, 2022). Other typical procedures, such as the Freiburg Acuity Test (Bach et al., 1996), rely on participants providing a response through button presses while visual stimuli of parametrically adjusted sizes are presented. In both of these cases, participants need to respond – either verbally or by pressing buttons.

Such examination is markedly more challenging to conduct with very young participants. Newborns or infants are not able to provide the required responses, cannot follow instructions of a complex psychophysical task, and are often inattentive or even asleep. Despite these difficulties, several methods have been established to estimate the visual capabilities of newborns and infants. While some of these include optokinetic nystagmus (Gardner and Weitzman, 1967), preferential reaching (e.g., Granrud et al., 1985; Yonas et al., 1982), and many others, I will focus here on reviewing two of the most commonly used methods for measuring the visual proficiencies discussed in 1.2.3. This includes behavioral methods based on preferential-looking behavior as well as electrophysiological methods based on visually evoked potentials.

1.2.2.1 *Behavioral examinations based on preferential looking*

One of the most commonly used techniques for assessing an infant's vision is the method of preferential looking, which was first invented by Fantz (Fantz, 1958, 1965; Fantz et al., 1962) and further developed to the method of forced-choice preferential looking by Teller (1979). What underlies the technique by Fantz (1958) is the observation that infants are more likely to fixate on visual stimuli that are 'interesting', in the sense that they exhibit some detectable patterns, rather than stimuli that are homogeneous and plane.

As described in Daw (2014), in a prototypical preferential looking experiment, an infant is positioned in front of two displays, and their attention is directed toward the displays. Then, an 'interesting' stimulus is presented on one of the two displays (either the left or the right one). For instance, if visual acuity is to be measured, the stimulus would be an acuity grating comprising patterns of alternating black and white lines, with the spacing between them corresponding to a particular spatial frequency. Across trials, the location of the 'interesting' stimulus would randomly vary between the left and right display, and responses to different stimuli – corresponding to different spatial frequencies – would be collected. The rationale behind this approach is that if the infant's acuity is high enough to resolve a given spatial pattern, the presented pattern would be rendered more 'interesting' than the neutral display, and the infant would be more likely to fixate on it. If, however, the spatial frequencies of a given pattern were too high to be resolved, then the infant, no longer able to detect the pattern, would be equally likely to attend to either of the two displays.

Through the further advancement of this procedure by Teller (1979), the preferential looking technique became a forced-choice preferential-looking technique. The forced-choice component thereby does not refer to the infant having to make a forced decision but to the experimenter who needs to do so. Specifically, the experimenter, not informed about which stimulus is presented on which display, needs to judge which of the two displays was more strongly attended in a given trial. The experimenter thereby takes into account the infant's head and eye movements. This procedure is repeated for several trials for each condition to be tested (e.g., a set of acuity gratings with different spatial frequencies). If, for a given condition, the experimenter judged that the 'interesting' display was attended to more often than expected by chance (typically, this threshold is set to 70% or 75%), it can be inferred that the stimulus was detectable for the infant. Following this basic methodology, the smallest detectable stimulus can be determined. In the case of visual acuity, this would refer to the highest spatial frequency with which the generated gratings were still detectable and attended; in the case of color vision, it would refer to the color with the slightest hue that was still sufficiently more 'interesting' than a purely gray pattern.

Over time, small variations of this technique have been proposed, in particular, to increase the speed of the experiment. In the case of acuity, this includes differences in specific psychophysical details (Atkinson et al., 1986) and the use of a simple acuity card procedure that enables assessments within just a few minutes (McDonald et al., 1985; McDonald et al., 1986). However, the method's fundamental rationale and approach remained the same and now constitutes one of the most widely used tests of infant vision (Daw, 2014).

1.2.2.2 *Electrophysiological examinations based on visually evoked potentials*

The preferential looking technique introduced in the previous section presupposes that the presentation of detectable visual patterns induces an infant to fixate on such patterns preferentially. Examinations based on visually evoked potentials (VEPs), instead, are based on the idea that the presentation of such patterns elicits a brain response that can be reliably picked up using EEG (electroencephalography). In both cases, the stimulus dimension of interest can be varied parametrically, and the cut-off point corresponding to the minimal detectable stimulus can be determined. In the following, I will describe this approach's rationale and procedures based on the excellent accounts provided by Norcia et al. (2015) and Sokol (1976).

In a classic VEP-based experiment, an infant is directed to look at stimuli that are presented on a screen, while being connected to an EEG system, with electrodes picking up electrical potentials from the infant's scalp. It is important to note that event-related potentials (ERPs), which are the measured brain responses to a specific event, such as the presentation of a visual stimulus, cannot only be measured in response to an isolated event but also for stimuli presented periodically at a fixed frequency. These responses, termed steady-state visually evoked potentials (or SSVEPs), are oscillatory responses to stimuli presented periodically, with the frequency of the electrical responses matching that of the visual stimulation (Norcia et al., 2015; Sokol, 1976).

As reviewed in Norcia et al. (2015), one of the most typical use cases of SSVEP stimulation is the so-called sweep VEP (Regan, 1973), in which the SSVEP is being recorded while a stimulus dimension of interest is parametrically varied (or 'swept') over a range of relevant values. For instance, when interested in measuring a person's visual acuity, the spatial frequency of a certain visual stimulus can be varied systematically throughout the experiment. Subsequently, the resulting SSVEP data can be analyzed as a function of the spatial frequency. Considering the periodicity of the SSVEP, the recorded data are classically analyzed in the frequency domain. There, in case a presented pattern was detectable for the infant, the signals would be expected to exhibit narrow peaks at the stimulus frequencies as well as its harmonics (Norcia et al., 2015). With such analysis, one can extract the amplitudes of the peaks at such frequencies and depict them as a function of the varied stimulus parameter. Consequently, one can estimate at which point (e.g., from which spatial frequency on) there are no detectable brain responses to the visual stimulation anymore.

Continuing with the example of measuring visual acuity, while many different types of stimuli could be presented, the most common types of pattern stimuli are sine-wave gratings (alternating black and white lines) and checkerboard patterns (alternating black and white squares arranged in a two-dimensional grid) (Zheng et al., 2020). In the spirit of the SSVEP paradigm, these patterns are presented periodically at a given temporal frequency – either in an onset-offset mode (where they are periodically presented with a blank screen in between) or, more commonly, in a pattern-reversal mode (where the bright and dark parts of the stimulus alternate back and forth) (Zheng et al., 2020).

Overall, this technique has been established for many measurements, such as refractive errors (Regan, 1973), visual acuity (Tyler et al., 1979), contrast sensitivity (Allen et al., 1986; Norcia et al., 1985), or color (Allen et al., 1993; Kelly et al., 1997; Ver Hoeve et al., 1996). It has since been used widely and it contributed significantly to characterizing visual proficiencies not only in infants (see Section 1.2.4.) but also in clinical populations, such as patients with spastic cerebral palsy (Costa et al., 2004) or cortical visual impairment (Good, 2001).

1.2.3 *Developmental trajectories of visual proficiencies*

After having introduced two of the most commonly used techniques for assessing visual proficiencies in infants, I will next provide an overview of findings that have resulted from the use of these techniques. Considering the work presented in this thesis, I will focus on the dimensions of visual acuity and color sensitivity in particular.

1.2.3.1 *Visual acuity*

Many behavioral and electrophysiological studies have examined the development of human visual acuity longitudinally following birth. While behavioral examinations would examine a newborn's or an infant's preferential looking behavior in response to the presentation of a visual grating of a certain spatial frequency, EEG-based examinations would classically be based on a frequency analysis of signals recorded during the presentation of a periodically reversing visual grating or pattern with 'sweeping' spatial frequencies (Daw, 2014).

A finding that has emerged consistently across both techniques is the marked improvement of visual acuity from birth to about 6 months of age (Daw, 2014; Dobson and Teller, 1978; Dobson et al., 1978; Fantz et al., 1962; Hamer et al., 1989; Marg et al., 1976; Norcia and Tyler, 1985; Sireteanu, 2000; Sokol, 1976; Teller et al., 1974). However, the absolute level of acuity that is estimated differs between the preferential looking and the VEP-based studies. As depicted in Dobson and Teller (1978), VEP studies, such as Sokol (1976) and Marg et al. (1976), lead to estimates of visual acuity levels that are markedly higher than the corresponding behavioral estimates. As discussed in Dobson and Teller (1978), the differences between the behavioral and electrophysiological estimates may be methodological and could simply be the result of preferential looking studies employing stricter threshold criteria than the VEP-based equivalent. Similarly, differences in the spatial and temporal parameters of the stimuli could potentially account for some of the observed differences. Alternatively, instead of representing methodological differences, the two methods could potentially measure signatures of different aspects of this visual proficiency (Chandna, 1991).

While the reasons for these behavioral and electrophysiological deviations are not yet clearly established, the differences do not distract from the main finding that visual acuity improves markedly during the first 6 months of life. After this initially rapid development, visual acuity continues to improve for several years. As reviewed in Leat et al. (2009), several studies based on preferential looking behavior demonstrate adult-like acuity levels by approximately 6 years (e.g., Birch et al., 1983; Ellemberg et al., 1999; Mayer and Dobson, 1982) while other studies still found slight differences to fully-mature proficiencies at this age (e.g., Heersema and Vanhofvanduin, 1990; Neu and Sireteanu, 1997; Stiers et al., 2003). It is important to note that the overall developmental trajectory of visual acuity is remarkable, evolving from a Snellen fraction of 20/600 at birth to 20/20 when fully matured (as described earlier, a person with an acuity of 20/600 would be able to see at a distance of 20 feet what a person with normal eyesight would be able to see at 600 feet). This marked improvement over time is due to the maturation of the retina and optics of the eye as well as that of the visual cortex (Banks and Bennett, 1988; Daw, 2014; Kiorpes and Movshon, 2004; Yuodelis and Hendrickson, 1986).

1.2.3.2 *Color sensitivity*

Another common assessment of visual proficiency concerns the ability to perceive and discriminate colors. In order to do so, one needs to be able to differentiate between different wavelengths of incoming light. This ability is rendered possible by the presence of three different types of photoreceptors in the human retina (S-cones, M-cones, and L-cones), which are sensitive to short, medium, and long wavelengths, corresponding roughly to blue, green, and red colors (Teller, 1998).

The development of our sensitivity to chromatic information has long been of great interest to researchers, dating back at least to Darwin (1877). As reviewed in Teller (1998), while earlier studies often showed that infants were able to differentiate between objects of different colors, most of these studies could not rule out the possibility that discrimination occurred on the basis of different colors being perceived at different brightness levels. It took time until proper laboratory techniques were established, such as the forced-choice preferential looking paradigm applied to chromatic discrimination with controlled luminance (Peeles and Teller, 1975) and its VEP-based analog (e.g., Crognale, 2002; Suttle et al., 2002).

Such studies have consistently revealed that color experience is extremely limited at birth but develops rapidly over the first few months of life. More specifically, while M and L cones were shown to function at two months by the latest, S cones exhibit functionality slightly later (Adams and Courage, 2002; Bieber et al., 1998; Knoblauch et al., 1998; Suttle et al., 2002; Teller, 1998). As summarized in Kellman and Arterberry (2007), while there is almost no evidence for human hue discrimination before four weeks of age, basic color vision is approximately adult-like by four months, with some differences in specific color properties remaining throughout the first year of life or even longer (Crognale, 2002; Crognale et al., 1998).

Thus, the development of visual acuity and color sensitivity is generally comparable in terms of the overall trajectory from very degraded to fully proficient. However, color vision appears to plateau earlier in its development than visual acuity.

1.2.3.3 *Other visual proficiencies*

As reviewing the development of the full set of visual proficiencies would exceed the scope of this introduction, the interested reader is referred to the excellent reviews by Daw (2014) and Kellman and Arterberry (2007). Overall, these developmental trajectories share similar tendencies, with the specific timelines and maturational mechanisms differing between them. Further, as illustrated in Ayzenberg and Behrmann (2022), there may be a complexity gradient in the maturational timeline, where basic proficiencies such as acuity or orientation selectivity develop earlier than more complex ones, such as the acquisition of viewpoint invariance or object categorization.

1.2.4 *The development of temporal frequency sensitivity in the auditory system*

Unlike in the visual domain, where birth induces the sudden onset of sight, significant auditory experience already begins prenatally. This observation is based not only on anecdotal reports of pregnant mothers who have often described movements of their unborn child in the presence of loud sounds from the external environment (Murkoff, 2016), but it is also based on systematic psychophysical experimentation, as detailed below.

Considering that auditory experience already starts prenatally, perceptual proficiencies cannot be examined using the classic behavioral or electrophysiological methods used for examining postnatal development. However, movements exhibited in response to environmental sounds can be detected already in the womb. Following this approach, in a remarkable study, Hepper and Shahidullah (1994) examined a total of 450 fetuses at gestational ages ranging from 19 to 35 weeks. In their experiment, the fetuses were exposed to pure sine waves (with frequencies of either 0.1, 0.25, 0.5, 1, or 3 kHz) through a speaker placed on the pregnant mother's abdomen. While these sounds were played, ultrasound videos were recorded. These were later presented to neutral observers, not informed about the specific condition tested in a given trial, who had to judge whether or not the fetus was exhibiting a response in a given trial. Following this methodology, this study revealed that at around 27 weeks of gestational age, the majority of fetuses responded to low frequencies (up to 500 Hz) but that none of them responded to the higher-frequency sounds (at 1 kHz and 3 kHz).

This finding indicates that the fluids and tissues in the intrauterine environment, along with other factors, markedly reduce the audibility of high frequencies, rendering the prenatal experience of a fetus effectively a low-pass filtered one (Hepper and Shahidullah, 1994). This result is in line with studies of pregnant sheep, in which externally presented speech was recorded from the inner ear of a fetal sheep and later presented back to human listeners (Gerhardt and Abrams, 1996; Griffiths et al., 1994; Smith et al., 2003).

To conclude, unlike in the case of visual acuity, where a newborn is exposed to blurry inputs at birth and experiences more and more high-resolution inputs later on, in the auditory domain, this transition happens from the prenatal period (where the intrauterine environment induces a strong low-pass filtering) to the postnatal period (outside of the intrauterine environment, with quite mature auditory frequency sensitivity immediately). Thus, while the timelines and mechanisms differ strongly between the two modalities, there is a commonality between them. In Section 1.5, I will further elaborate on the potential significance of this commonality in the context of the AID hypothesis.

1.3 ATYPICAL PERCEPTUAL DEVELOPMENT

Studies of perceptual development, as reviewed in the previous section, have led to the detailed characterization of typical developmental trajectories. However, there are millions of children worldwide whose perceptual experience deviates from normal development. These include, among many others, individuals who are born blind and those who experience sensory processing deficits following premature births. In this section, I will motivate the work with atypically developed individuals from two perspectives. First, carefully conducted examinations of such individuals are critical to help develop more accurate clinical prognoses and to aid the design of appropriate rehabilitative interventions. In addition, such studies also have the potential to raise awareness of the practical treatability of certain developmental conditions along with their societal need. Second, in addition to their direct humanitarian impact, studies of individuals who experienced atypical trajectories of sensory development can also meaningfully add to our scientific understanding of normal development. This is because individuals who experience deviations from normal development provide a rare but unique opportunity to examine the causal consequences of such deviations. Such examinations, in turn, can shed new light on the significance of key characteristics of normal development.

1.3.1 *The case of treatable childhood blindness*

Considering the work presented in this thesis, I will focus here on introducing a specific group of atypically-developed children – those who were born blind and received treatment for their blindness late in life. I will begin by reviewing the societal need for, as well as the medical feasibility of, treating congenitally blind children in the developing world. I will then motivate the unique scientific opportunities that come with the post-surgical examination of such children. Finally, I will describe and discuss past efforts that were undertaken as part of 'Project Prakash' ('Prakash' as the Sanskrit word for 'light') (Mandavilli, 2006; Sinha, 2013). This initiative, which I have had the great pleasure of working with, aims to combine the humanitarian and scientific missions of providing free sight surgeries to curably blind children in India and studying their visual development following late sight-onset.

1.3.2 *The societal need for treating curably blind children*

The World Health Organization has estimated that more than 1 million children worldwide suffer from blindness (Gilbert and Awan, 2003; Gilbert and Foster, 2001; World Health Organization, 2000). Most of these children lead immensely deprived lives; many even die. This is especially true for the approximately 75% of blind children who are born in developing countries (Steinkuller et al., 1999). Gilbert and Awan (2003) attributed the markedly greater proportion of blind children in the developing world to three main factors: the greater presence of potentially blindness-inducing conditions, such as vitamin A deficiency; the lack of control over conditions such as the measles; and the lack of access to medical centers for carrying out appropriate eye surgeries. For blind children born in developing countries, the mortality rate in the years following the diagnosis is estimated to be above 50% – about five times as high as in developed countries (Gilbert and Foster, 2001). Those who survive are likely to lead remarkably difficult lives on the social, personal, educational, and economic front (Gilbert and Foster, 2001; Rahi et al., 1999).

In light of the above, it is important to point out how much of this suffering could be lessened. While the specific numbers vary between factors such as geographic regions, it is generally estimated that in more than 25% of these cases, blindness could have been treated or prevented (Bhalerao et al., 2015; Bhattacharjee et al., 2008; Gilbert et al., 1993; Gilbert and Awan, 2003; Steinkuller et al., 1999). Thus, while some conditions of visual impairment cannot, as of yet, be prevented or treated, other conditions, like the presence of cataracts or corneal opacities, are easily treatable and make up a significant portion of the cases of childhood blindness in the developing world (Khanna, 2018; Steinkuller et al., 1999). Considering the medical feasibility of providing sight-restoring surgeries to children suffering from these forms of blindness, doing so can help meet a grand societal need – even if post-surgical improvements were only partial.

1.3.3 *The scientific opportunity of examining visual development following treatment for congenital blindness*

In addition to meeting a great humanitarian need, treating curably blind children and subsequently examining the development of their visual proficiencies also provides unique scientific opportunities.

First, examining the resilience of perceptual proficiencies to long periods of visual deprivation immediately following birth not only contributes to better prognoses for the effectiveness of treating congenital blindness, but it can also advance our current understanding of brain plasticity and how the sensory system can learn to reorganize late in life. Historically, much of the work on the resilience to early-onset visual deprivation has been conducted with animals. These studies – notably, Wiesel and Hubel (1965) with kittens and Hubel et al. (1977) and Le Vay et al. (1980) with macaque monkeys – revealed markedly limited capabilities of acquiring normal vision following extended periods of initial deprivation. These findings gave rise to the notion of 'critical periods' or 'sensitive periods' of brain development (for a review, see Kiorpes, 2015). This describes the idea that perceptual development critically relies on access to appropriate sensory inputs early in life when neural plasticity is still high. However, as findings from animal studies cannot be directly applied to humans, working with the rare population of children treated for blindness late in life (henceforth also referred to as late-sighted individuals) can significantly contribute to our understanding of the timeline and mechanisms of plasticity in the human brain. Further, when working with animals, examinations of perceptual proficiency are restricted to analyzing physiological signals or simple behavioral responses. Working with children or young adults, in contrast, allows for the analysis of more complex behavior, which is crucial for assessing many higher-level perceptual abilities.

Second, working with late-sighted individuals allows for examining a wide range of visual proficiencies with complex behavioral and neuroimaging techniques, immediately and longitudinally following the sight-restoring surgeries. Such complex examinations of the very initial stages of visual development would be unimaginable to be carried out with newborns. This is, as reviewed in Section 1.2., due to the operational difficulty and limitations of experimental methodologies. Thus, studies on late-sighted individuals can provide unprecedented insights into complex perceptual function at the very initial stages of visual development. It is important to acknowledge, however, that given the differences between a newborn and a newly-sighted child, the initial developmental stages in the late-sighted should not be mistaken as necessarily being equal to those in a normally developed child.

Third, the above caveat concerning inevitable differences between a newborn and a newly-sighted child, while important to acknowledge, can, in fact, also be seen as an opportunity. Specifically, newborns and late-sighted individuals differ not only in terms of their experience prior to sight onset but also in terms of their experience following it. For instance, some of the maturational processes that normally take place in the first months or years following birth, also proceed in a child that is suffering from blindness. As a consequence, the visual system of a child whose vision had just been restored is markedly more mature than that of a neonate, for instance in the domain of visual acuity (Boas et al., 1969; Hendrickson and Boothe, 1976). This provides an opportunity to examine the consequences of unusually mature initial visual experience in late-sighted children and, in turn, the potential significance of initially degraded experience in normally-sighted newborns. This perspective, central to the work presented in this thesis, will be further elaborated on in Section 1.5.

1.3.4 *Project Prakash: merging science and service*

In the previous sections, I have highlighted the societal significance of providing treatably blind children with sight surgeries, along with the unique scientific opportunities of examining the children's visual development immediately and longitudinally following their sight onset. Project Prakash (Mandavilli, 2006; Sinha, 2013; Sinha and Held, 2012) is a joint humanitarian and scientific initiative, launched in 2003 and located in Delhi, India, which aims to achieve exactly these goals. To this end, Project Prakash is comprised of three main components – outreach, treatment, and research. In the following paragraphs, I will summarize its logistics and operations, based on what has been reported in several publications (Ganesh et al., 2014; Sinha, 2013, 2016; Sinha et al., 2013).

The first and logistically most complex component of Project Prakash comprises rural outreach. As part of this, outreach team members conduct eye screening camps in schools for the blind and other communities across numerous villages in northern India, where medical access is typically scarce. Should the initial screening indicate that some of the children could benefit from eye surgeries, they are arranged to be brought for detailed ophthalmological examinations to Dr. Shroff's Charity Eye Hospital in Delhi – the medical partner of Project Prakash. Such examinations are useful if a child exhibits signs of an occlusive eye pathology like a cataract, which can easily be removed, as opposed to suffering from a permanent eye condition for which no treatment is available yet. Should the follow-up investigation in the hospital confirm the usefulness of surgical treatment, the child and parents are informed about the possibility of receiving sight-restoring surgeries free of cost. These surgeries typically include both the removal of the cataract and the insertion of an intra-ocular lens.

The data reported in Ganesh et al. (2014) illustrate the complexity of the logistics involved in Project Prakash. As detailed in the paper, over a period of four years, more than 20,000 children have been examined as part of the eye screening camps, more than 1,000 were referred to the hospital, 427 children were subsequently advised to receive surgical treatment, and the families of 237 individuals agreed to the surgeries being carried out. While all of these children were then, indeed, provided surgeries, for the scientific study reported in the paper, only 53 of the treated children met the specific inclusion criteria, such as having congenital or very early-onset blindness and exhibiting severe occlusion (Ganesh et al., 2014). The latter is illustrative of how meeting the specific inclusion criteria for the planned scientific studies is not a precondition for receiving medical treatment free of cost.

From the initiation of Project Prakash in 2003 to 2016, more than 40,000 children's eyes have been screened, with over 450 children having been provided surgeries and almost 1,500 children having received non-surgical treatment (Sinha, 2016). These numbers continue to increase steadily each year. While Project Prakash is, thus far, focused on conducting eye screening camps across northern India – motivated by India being home to the largest population of blind children (Dandona and Dandona, 2003; Sinha, 2013) – similar initiatives have recently started to emerge in other developing countries (for the work resulting from an effort in Ethiopia, see, e.g., Senna et al., 2021), indicative of a more global movement.

Following eye surgeries conducted at Dr. Shroff's Charity Eye Hospital, the Prakash children's vision is being examined. The children are also invited to participate in longitudinal scientific studies following the surgeries, but this is non-obligatory for any of the children.

1.3.5 *Past findings that have emerged from the work with late-sighted individuals*

Scientific studies conducted with late-sighted children from Project Prakash and other initiatives have led to several important insights. Perhaps most important of these is the consistent and marked improvement in basic visual proficiencies, such as visual acuity, that is observed longitudinally following treatment for congenital blindness (Ganesh et al., 2014; Kalia et al., 2014; Mandavilli, 2006; Ostrovsky et al., 2006; Sinha, 2013; Sinha and Held, 2012). Even though the late-sighted typically do not fully reach the level of normally-sighted controls, this finding is remarkable, considering the resulting improvement in life quality (Kalia et al., 2017) as well as the scientific implications for our understanding of 'critical periods'.

Most of the previous animal studies that examined resilience to induced visual deprivation at birth have demonstrated the substantial limitations of acquiring typical visual proficiencies if deprivation lasted longer than what has been proposed to be a 'critical period' of development (Hubel et al., 1977; Le Vay et al., 1980; Wiesel and Hubel, 1965). Considering the advanced ages of late-sighted individuals (with mean ages typically greater than ten years in Project Prakash), the results emerging from the corresponding human studies argue against this notion in the strict sense. As discussed in Sinha and Held (2012), the human and animal findings are thereby not technically contradictory as there are significant technical differences between the two. For instance, in addition to the examinations having been carried out in human vs. non-human species, many of the deprivation experiments in animals have been conducted monocularly. In contrast, the visual impairments observed in human patients are binocular ones. Further, the set of animal experiments that did induce binocular deprivation have typically done so in a much more extreme way than what would correspond to blindness induced by a cataract, as is typically the case in a human patient (Sinha and Held, 2012). Thus, while one needs to be careful in making direct cross-species and cross-method comparisons, the basic patterns of results that emerge from the studies with late-sighted humans speak for the resilience to visual deprivation and significant remaining plasticity, questioning the notion of critical periods in the strict sense.

Further, while past animal experiments were restricted to examining the resilience of relatively low-level visual skills to deprivation, working with late-sighted patients allows for complementing these with a broader set of more complex, high-level visual proficiencies. Past examinations of such proficiencies include, among others, categorical face perception (Gandhi et al., 2017), susceptibility to visual illusions (Gandhi et al., 2015), as well as temporal processing and the use of motion cues (Ostrovsky et al., 2009; Ye et al., 2021). In most of these cases, while not fully reaching normally-sighted controls, Prakash patients exhibit marked improvements immediately or longitudinally following the surgeries. Along similar lines, studies from both Project Prakash and the initiative in Ethiopia have revealed that the cross-modal mapping between vision and touch is not established immediately post-surgery but does develop with experience thereafter (Held et al., 2011; Senna et al., 2021).

However, while many visual proficiencies markedly improve following the surgeries, there appear to be some proficiencies that do not. For instance, Piller et al. (2023) reported that grasping behavior does not recover post-surgically. Also, in different populations of late-sighted children, the ability to identify faces remained impaired (Geldart et al., 2002; Vogelsang et al., 2018). As will be further elaborated in Section 1.5., these specific impairments provide an opportunity to study their potential origin. Better understanding such origin could, in the long run, give rise to the design of rehabilitative interventions as well as a better understanding of the corresponding normal developmental processes.

1.4 deep neural networks

After having motivated and presented the study of typical and atypical perceptual development above, I will now turn to introducing deep neural networks. These networks will later serve as computational model systems to answer questions about the role of early sensory experience for the establishment of later perceptual mechanisms. Before discussing deep networks in the specific developmental context that is key to this thesis (see Section 1.5.3.1), I will, in this section, describe and evaluate them as models of the human perceptual system more broadly.

1.4.1 *The motivation for using deep neural networks as computational models*

Many possible motivations exist for using deep neural networks as computational model systems. From the engineering perspective, these networks represent the state-of-the-art for solving complex tasks in several domains, such as vision or language. Working towards further advancing these technologies can provide significant value for the research community as well as the general public. From the scientific perspective, deep neural networks can serve as excellent tools for testing precisely formulated scientific hypotheses, particularly (though not exclusively) if they can provide data that would be challenging or infeasible to collect in humans or animals. As I will elaborate on further below, key to their utility is their ability to learn directly from input data as well as their scale, allowing for comparisons of performance on perceptual tasks that humans can also be examined on.

In this section, I will begin by providing a very brief overview of the history of artificial neural networks (ANNs) – from the development of McCulloch & Pitts neurons in 1943 to modern-day deep networks capable of performing complex classification tasks. I will then describe modern-day convolutional neural networks and examine the utility of using these networks as models of the human perceptual system, based on structural, behavioral, and representational considerations. It is important to note that the review of this section – and the work presented in this thesis – is restricted to deep neural networks (primarily, in the visual domain) and does not include more biologically detailed models.

1.4.2 *A brief history of artificial neural networks*

The effort of modeling neurons and networks of neurons has a long and somewhat erratic history, dating back to at least the first half of the twentieth century. It was in 1943 that, inspired by the discrete nature of neural activity in the brain, McCulloch and Pitts (1943) developed a theoretical model of an artificial neuron. This model was presented as a 'logical calculus', where a neuron operated by receiving binary inputs and computing a binary output based on a simple threshold function. A few years later, Hebb (1949) introduced the critical concept of Hebbian learning, commonly known under the phrase "what wires together, fires together". This phrase refers to a learning procedure in which the connections between neurons that are simultaneously active get strengthened over time, thereby providing a candidate implementation of learning in neural networks.

Following these important early contributions, a big step forward in the establishment of ANNs was made about a decade later with the invention of the 'Perceptron' by Rosenblatt (1958). The Perceptron was a simple single-layered neural network, implemented as a physical machine. The demonstration that this machine could perform the first simple visual classification tasks elicited great enthusiasm in the field. However, about ten years later, Minsky and Papert (1969) published their criticism that single-layer perceptrons are necessarily restricted in their computational capabilities as they can only solve a very restricted type of classification task (specifically, classification for linearly separable classes). This insight led to a marked reduction in the enthusiasm for the neural network modeling endeavor at that time.

A key driving force behind the further development of new neural network architectures was the groundbreaking neurophysiological work by Hubel and Wiesel, who managed to carry out single-cell recordings of the cat visual cortex. With their recording techniques, they were able to identify different populations of neurons – so-called simple cells and complex cells – and gave rise to the idea of the visual cortex being organized hierarchically (Hubel and Wiesel, 1959, 1962). These findings inspired Fukushima (1980) to develop the 'Neocognitron'. This is a neural network model comprised of multiple connected layers following a hierarchical organization scheme, which also incorporated the notion of local receptive fields. While it was clever in its design and laid the foundation for convolutional neural networks later on, the technical limitations at the time greatly restricted its practical utility. Before neural networks could ultimately demonstrate their ability to solve complex visual classification tasks, several technological advancements had to first come by.

A particularly crucial one was the invention of the backpropagation algorithm by Rumelhart et al. (1986), which allowed for an elegant way of updating the weights of a multi-layer neural network during training. This technique enabled LeCun et al. (1989) to train a 4-layer convolutional neural network (CNN) to recognize handwritten digits. However, additional advancements were needed before convolutional neural networks would be positioned at the forefront of computer vision technologies. Eventually, this occurred with the emergence of large datasets of labeled images, such as the ImageNet database (Deng et al., 2009), the availability of fast graphics processing units that were used to carry out complex neural network computations, and further refinements of network architectures as well as extensions of their depth. These factors together enabled Krizhevsky et al. (2012) to develop the AlexNet – the first neural network to win the ImageNet visual recognition challenge in 2012 by significantly outperforming all other models. Since then, deep neural networks have been immensely successful in the domain of engineering, also fueled by the development of more complex architectures (e.g., ResNet (He et al., 2016) and Inception (Szegedy et al., 2015)) as well as the incorporation of additional features.

Before examining the suitability of deep neural networks as models of the human perceptual system, I will first provide some insight into how these networks actually work. While, in the context of this thesis introduction, this account must be fairly concise and rather conceptual, readers interested in the mathematical foundations of deep learning are referred to the excellent books by Goodfellow et al. (2016) and Bishop and Nasrabadi (2006), on which the below is partially based.

1.4.3 *A brief introduction to deep convolutional neural networks*

In this sub-section, I will describe the functionality of deep convolutional neural networks (DCNNs) using the example of the AlexNet (Krizhevsky et al., 2012). This is a moderately deep and relatively simple convolutional neural network with a total of eight network layers (five convolutional layers followed by three fully-connected ones). As the individual components of the AlexNet are relatively typical, much of this explanation also applies to other DCNNs.

I will begin by describing how a trained network is capable of transforming an input (a given image) into an output (a classification decision; for instance, the predicted object class of the provided input image) as part of what is termed the 'forward pass'. Next, I will explain how DCNNs are able to learn these associations through training, by drawing on the 'forward pass' as well as the 'backward pass'. Finally, I will mention several considerations to be made when training a DCNN and will briefly describe how a network can be analyzed following training.

1.4.3.1 *The forward pass: from input image to classification decision*

As the first step of the forward pass, an image (sized 227 x 227 x 3 pixels, with the last dimension corresponding to the RGB channels) is fed into the input layer of the AlexNet (Krizhevsky et al., 2012). This input image, stored in terms of its RGB values for each pixel, is typically normalized to enable the distribution of individual pixel values to have a mean of zero and a standard deviation of one, for each of the RGB channels. (Note that, for reasons detailed in Chapters 3 and 4, in the studies reported in this thesis, I often simply scaled the input from a range of $[0, 255]$ to a range of $[-1, 1]$.)

In the first convolutional layer of the AlexNet, the (rescaled/normalized) input image is then convolved with a set of 96 kernels (also referred to as filters or receptive fields), each being sized 11x11x3 pixels (Krizhevsky et al., 2012). As part of the convolution operation, each kernel is systematically slid over the input image, resulting in 96 two-dimensional feature maps representing the similarity between each kernel and different parts of the input image. As the kernels in a trained network have been learned to take on specific shapes, such as Gabor-like patches, important visual features, such as edges, can be extracted by convolving the input image with the network's kernels (Goodfellow et al., 2016).

Next, an activation function, responsible for rendering the DCNN responses non-linear, is applied to the feature maps that resulted from the convolution operations. While there are several types of activation functions, the most common one, which has also been used in the original AlexNet paper (Krizhevsky et al., 2012), is the Rectified Linear Unit (ReLU) function: $f(x) = max(0, x)$. This function simply translates all negative activations to zero.

Then, the max-pooling operation is applied to the ReLU-transformed feature maps. This serves as a simple dimensionality reduction technique by taking the maximum value within a certain neighborhood (usually, 2x2 pixels) of the input. Typically, a procedure called batch normalization is applied to the outputs of the convolutional layers prior to the application of the max-pooling operation, in order to ensure that the max-pooling layer receives normalized inputs. It is worth noting, however, that the original AlexNet did not yet incorporate batch normalization, as this technique was only introduced later (Ioffe and Szegedy, 2015).

The outputs of the max-pooling layer subsequently serve as the inputs to the second convolutional layer. The process is then repeated all the way until reaching the fifth and final convolutional layer. The only difference between the layers is that only the first, second, and fifth convolutional layers are followed by the max-pooling operation.

The output of the fifth convolutional layer then becomes the input to the first fully-connected layer. Fully-connected layers thereby do not perform the convolution operation but compute, in essence, a weighted sum of the features from the previous layer. As the weights of this weighted sum are learned, fully-connected layers are able to combine complex combinations of features extracted in the previous layers, in order to arrive at a classification decision. The output of the first fully-connected layer subsequently becomes the input to the second fully-connected layer, which, after being processed in the second fully-connected layer, becomes the input to the third, and final, fully-connected layer.

Three special properties within the fully-connected layers should be noted. First, the first two fully-connected layers are each followed by dropout layers (Hinton et al., 2012). In these layers, the activation of each unit is set to zero with a certain probability (here, 50%). This forces the network to learn based on a wider range of features, thereby rendering it more robust. Second, the final fully-connected layer is equipped with as many units as there are different classes in the dataset to be classified. Third, the softmax operation is applied to the output of the final fully-connected layer to transform the values of this layer into a distribution representing the probability for a given input image to belong to each of the classes that are part of the dataset. This distribution is then transformed into a discrete classification decision by choosing the class label corresponding to the highest probability.

This concludes the forward pass, going all the way from an input image to a discrete classification decision. Next, I will explain how this matching is actually achieved as part of the network training procedure.

1.4.3.2 *The backward pass: from classification errors to adjusted weights*

A typical dataset used for training DCNNs is the ImageNet database (Deng et al., 2009), comprising more than one million images belonging to one thousand different object categories. Each image is thereby associated with one of the thousand labels, and the network is trained on classifying each image into one of the thousand classes. As the labels act as a teaching signal, this approach belongs to the broader class of supervised learning algorithms.

Prior to training, all the weights of the network are initiated randomly. This means that the kernels in the convolutional layers do not yet exhibit any 'useful' structures that would help extract critical visual features. Further, the fully-connected layers are not yet equipped with 'useful' weights to compute complex weightings of previously extracted features.

When beginning the training process, images from the training set are sent through the network as part of the forward pass. This results in classification decisions for each of the images. Initially, the correctness of these classifications is at chance level. To quantify the discrepancy between the predictions of the network and the correct labels (which are provided for each individual training image), a loss function is applied. While there are different types of loss functions, the Sparse Categorical Cross Entropy loss function is typically used when dealing with discrete class labels, such as in the case of the ImageNet dataset.

After the loss is computed, which penalizes incorrect predictions, the backpropagation algorithm (Rumelhart et al., 1986) is applied. This is an algorithm used to compute the gradient of the loss function concerning each learnable weight in the network. Specifically, this algorithm works by computing the gradient at the output layer of the network and then distributing it backward through the layers, as part of the 'backward pass'. Over time, through this 'learning signal', the weights become less random and begin to enable more useful computations.

Important for the iterative adjustment of model parameters, focussed on minimizing the loss, is the use of an optimization algorithm. While there are many different algorithms, stochastic gradient descent (SGD) is one of the most common ones (Goodfellow et al., 2016). An important parameter for this optimization is the learning rate, which controls the step size of adjustments made in response to the error signal that is being backpropagated through the network. This parameter is especially important for considerations of convergence: too low learning rates converge slowly or could lead to the network only finding local minima; too high learning rates, on the contrary, may overshoot the optimal point in the parameter space.

With this machinery in place, the weights of the network are iteratively updated according to the gradient of the loss function, in order to minimize the discrepancy between the predicted classification decisions and the true labels of the training data. Over time, specific spatial structures emerge in the kernels of the convolutional layers, which serve to extract critical features of the images. Then, through the weights learned in the fully-connected layers, such extracted features are combined in a complex manner to arrive at a classification decision.

1.4.3.3 *Typical considerations to avoid overfitting*

When training a DCNN, a few considerations need to be made in order to prevent overfitting. This is a phenomenon where a model has been optimized too heavily on the training data and is not only fitting its 'signal' but also its 'noise'. As a consequence, the model may not grasp the essential features of the dataset, and classification abilities may not generalize well to images that were not presented during training. Such overfitting can be avoided, at least partially, by several means.

First, the number of epochs used for training needs to be selected carefully. This number thereby refers to the number of times that the entire training set is fed into the network and used to adjust the weights through backpropagation. A classic approach for selecting the number of epochs is based on tracking the loss and accuracy on the training set as well as that of a validation set (comprising images that the network has not been exposed to during training). If one trains for too few epochs, the regularities in the data have not been sufficiently incorporated into the network, and the training and validation accuracies are not high enough. If one trains the network for too long, overfitting may occur, in which case the classification accuracy on the validation and test set may begin to decrease.

Another important technique to avoid overfitting is based on data augmentation. As part of data augmentation, certain changes or perturbations are added to the input image, to enable the model to learn to classify images regardless of the perturbations. While there are many data augmentation techniques, I used only two for most papers reported in this thesis. These include the random cropping of an input image (from an original image of 256 x 256 pixels to a cropped 227 x 227 image, using different cropping windows) as well as randomly flipping the images horizontally (with a probability of 50%). Note that while there are many other techniques, such as the application of random blurring or adjustments of color or luminance information in the images, due to the nature of the papers I have reported in this thesis, I have not utilized those.

Finally, it is worth mentioning that the inclusion of a dropout layer (Hinton et al., 2012) in the network architecture can also help prevent overfitting. The rationale is that by a large proportion (in the case of the AlexNet, 50%) of unit activations being set to zero, the network is forced to learn a wider range of features that can be relied on.

1.4.3.4 *Analyses of neural networks following training*

After a network has been trained, one can examine its performance on a specific test set. Further, one can make modifications to the test set to study how generalized the performance profile is. To examine the inner workings of the network, one can also depict the learned filters of the first convolutional layers, as these typically depict interpretable shapes that bear at least superficial resemblance to receptive fields found in the cortex. In addition, one can also analyze the activity patterns in deeper network layers, for instance, by synthesizing artificial images that elicit maximal activity in a given unit of a given network layer. With these tools, and many more, one can train a network in different manners (for instance, in a developmental and a non-developmental way; see Section 1.5. for more details) and examine the consequences of these differences.

This concludes my short introduction to deep convolutional neural networks. I will now discuss how suitable these networks are as models of the human perceptual system.

1.4.4 *Deep convolutional networks as models of the human visual system*

While deep neural networks have emerged as excellent engineering tools capable of solving complex classification tasks with high accuracy, they have also played a crucial role in the scientific domain – as models of the human perceptual system. Thus, I will next discuss their suitability for the latter, and will do so based on structural, performance-based, and representational considerations. Note that this discussion is primarily based on deep convolutional neural networks operating in the visual domain but is similar to those operating on other inputs.

1.4.4.1 *Structural considerations*

Before diving into an assessment of deep neural networks as models of the human visual system on the basis of behavioral and representational comparisons, a general consideration has to be made regarding the desired level of biological detail vs. abstraction. Several design features of deep (convolutional) neural networks – such as their hierarchical structure, local receptive fields, convolutional layers, and their general organization as a network of connected units – were inspired by neuroscience. Nevertheless, deep neural networks significantly abstract away from most of the details present in the biological brain, both at the level of single neurons (such as their low-level molecular processes) as well as that of networks of neurons (concerning, for instance, the complexity of neural connectivity patterns in the biological system). Further, the standard approach in deep learning of carrying out supervised training with labeled input data, followed by weight adjustment with the backpropagation algorithm, does not match the human learning experience.

However, the lack of finer-grained biological detail in their implementation and procedures does not necessarily disqualify deep networks if they are to serve as a model, not a replica, of the biological system. To the contrary, if one were to implement each and every detail of the biological brain, it may be argued that the complexity of the resulting model, in addition to being too computationally costly, could distract from the underlying principles of brain organization. As elegantly phrased in Doerig et al. (2023), "ANNs live in the Goldilocks zone of biological abstraction (. . .), striking the required balance between biological realism and algorithmic clarity."

Furthermore, the lack of detail in current neural networks, together with some of the remaining shortcomings reviewed further below, might even provide an interesting scientific opportunity. Specifically, this motivates the introduction of additional features into deep neural networks while, in turn, systematically examining the consequences of such additions. Such an endeavor would be markedly more challenging, if not infeasible, to achieve in biological brains. This possibility relates to what Doerig et al. (2023) termed the "neuroconnectionist research cycle" – the idea that biological details can be added to a given network, which can then be evaluated and, based on the results, inform the creation of new models.

1.4.4.2 *Performance-based considerations*

In terms of their overall performance on large-scale image recognition tasks, beginning with the successes of the AlexNet (Krizhevsky et al., 2012), deep neural networks have demonstrated impressive capabilities, meeting or even exceeding what would be expected in humans. Thus, in contrast to most previous computational models, these networks offer the remarkable opportunity to examine them on challenging, high-level visual tasks that humans can also be tested on.

Despite this remarkable achievement and opportunity, it is important to note that empirical studies suggest that once the characteristics of the image set used for testing deep networks diverge from that of the training data, performance typically drops rapidly. For instance, Geirhos et al. (2018a) showed that deep networks are markedly less robust than humans to a vast range of image perturbations such as blur or noise. Similarly, deep networks are known to be sensitive to adversarial attacks (Szegedy et al., 2013) – these are systematic perturbations to images that are so small that they are not detected by humans but drastically reduce deep network classification performance. Along similar lines, deep networks have been shown to be more biased towards local texture than global shape in terms of their overall classification behavior and to engage in the learning of shortcuts, allowing for good performance on a certain benchmark but no deviations from it (Geirhos et al., 2021, 2018b). Thus, the question concerned with whether and how human-like generalization may be accomplished in deep networks is an important but mostly unresolved one.

1.4.4.3 *Representational considerations*

In addition to comparing deep networks and humans in terms of their classification performance on certain test sets and variations thereof, one can also evaluate the consistency between activation patterns in deep networks with those present in neural recordings of humans or animals. This effort, based on analyzing the representational similarity between activations in both systems (see Kriegeskorte et al., 2008), has recently been further facilitated by the availability of open benchmarks, such as those provided on the Brain-score platform (Schrimpf et al., 2020a,b).

This line of investigation has demonstrated that deep neural networks are the best models available for predicting neural activities (see, e.g., Cadena et al., 2019; Cadieu et al., 2014; Cichy et al., 2016; Khaligh-Razavi and Kriegeskorte, 2014; Storrs et al., 2021) as well as behavior (see, e.g., Kubilius et al., 2016). Nevertheless, the overall similarity scores computed between the computational and the biological system still have great potential for further improvement.

In addition to directly examining representational similarity between the activations of deep networks and the biological system, which would result in the extraction of a correlation-like score, deep networks also offer the possibility of visualizing their learned features. This allows for a comparison to those reported in neurophysiological studies, such as the receptive fields reported in the primary visual cortex (Hubel and Wiesel, 1959, 1962).

1.4.4.4 *Conclusion*

To conclude, while deep networks have been inspired by biological brains, they are far from being replicas of the latter. Further, while excellent at classifying large-scale visual datasets, deep network classification abilities do not usually generalize to variations in inputs that were not available during training. However, demonstrations of representational similarity in activation patterns between deep neural networks and the biological system render deep neural networks the best models of the human perceptual system that are currently available. In addition, the remaining shortcomings in generalization performance and some structural deviations provide a scientifically meaningful opportunity to refine deep network models further and carefully examine the consequences of these refinements. The latter approach has been exemplified by, for instance, the incorporation of recurrent connections (e.g., Kietzmann et al., 2019; Spoerer et al., 2017; Tang et al., 2018), the training on images devoid of local texture, thereby highlighting global features (Geirhos et al., 2018a), and the training on more ecological databases (e.g., Mehrer et al., 2021).

Another possibility for further advancing the design of deep learning training procedures is to incorporate knowledge about typical developmental trajectories. This approach will be further detailed in Section 1.5.

1.5 THE 'ADAPTIVE INITIAL DEGRADATION' (AID) HYPOTHESIS

1.5.1 *Introducing the hypothesis*

As reviewed in Section 1.2, many aspects of perceptual development follow a temporal trajectory from initially very degraded to fully proficient later on. While, in the visual domain, this trajectory unfolds following birth, in the auditory domain, it already begins to do so prenatally. These developmental progressions have generally been well-established. However, very little is known about their potential significance in helping to set up sensory processing strategies during early development.

In the more traditional view, the early limitations characteristic of normal development have typically been considered nothing but epiphenomena accompanying the physiological maturation of the developing sensory system. As such, they merely represent obstacles that must be overcome in order to achieve perceptual proficiency. A more recent proposal, however, has posited that, instead of being a hurdle, the developmental staging from limited to proficient may be adaptive and help, instead of hinder, the acquisition of later perceptual proficiencies. I refer to this as the 'Adaptive Initial Degradation' (AID) hypothesis.

The basic intuition behind this hypothesis is that by early experience being restricted to coarse-grained sensory inputs, the developing nervous system is forced to establish processing strategies capable of stably integrating such coarse-grained information. This restriction prevents becoming overly reliant on the presence of fine-grained details. For instance, in the case of visual acuity, the initial immaturities of the retina and cortex render the visual image so blurry that small segments of the essentially low-pass-filtered version of a visual scene do not provide enough information to be informative. As one possibility for overcoming this limitation, the development of spatial integration capabilities across larger parts of the visual scene could help to nevertheless derive meaningful inferences about the external environment (see Figure 5.1 in Chapter 5 for illustration). A similar story may be at play in the auditory domain. Although the development begins prenatally instead of postnatally and concerns temporal instead of spatial frequencies, the presence of exclusively low temporal frequencies in the period prior to birth may lead to the development of temporally extended integration mechanisms in order to derive meaning from the low-frequency structure of sounds. Similarly, in the domain of color vision, the lack of chromatic information at birth could induce the development of representations that are more tuned to luminance-based features and global shape information, which could prevent the learning of 'shortcuts', as part of which objects may be associated exclusively with color features.

In all of the above accounts, exposure to initially degraded sensory inputs would drive the development of perceptual mechanisms that are less reliant on low-level features. This development could provide important benefits for allowing more stable perceptual analysis later in life. This idea is at the core of the AID hypothesis. However, like all hypotheses, it needs to be carefully tested.

1.5.2 *Past work*

Some past work has presented first support for the idea that the initial degradations experienced as part of typical development may be beneficial rather than detrimental. Crucially, Turkewitz and Kenny (1982) laid out the theoretical argument that less complex stimuli at the onset of sensory experience can facilitate perceptual analysis and thereby support its acquisition. While these considerations were of a theoretical nature, Newport (1988) and Elman (1993) provided evidence for the benefits of such limitations – in the specific context of cognitive architectures and language learning. Elman (1993) thereby engaged in computational simulations with a simple recurrent neural network available at that time. In the sensory rather than the cognitive domain, the potential benefits of low visual acuity at the onset of sight have also been demonstrated in the context of learning to detect binocular disparities in a simple neural network model (Dominguez and Jacobs, 2003). More recently, the benefits of low initial acuity have also been shown by Vogelsang et al. (2018), who used the AlexNet CNN (Krizhevsky et al., 2012) and found that initial training on blurry faces, followed by training on high-resolution faces, resulted in the emergence of spatially-extended receptive field structures as well as more generalized face recognition performance (reviewed in greater detail in Chapter 5).

1.5.3 *Examining the AID hypothesis*

This past work motivates more comprehensive examinations of the role that initial degradations during early development may play in acquiring later perceptual skills across several perceptual dimensions and modalities. What especially motivates these examinations is that they would not only help contribute to our foundational understanding of typical development but might also have clinical significance for understanding perceptual deficits resulting from deviations from typical developmental trajectories, as happens, for instance, in Prakash children. Finally, insights gained from understanding normal development can inspire the more robust training of computational model systems.

In light of this significance, I have centered my thesis work around carrying out of such examinations (reported in Chapters 2-5). After having introduced and motivated the approaches of typical development (Section 1.2), atypical development (Section 1.3), and computational modeling (Section 1.4), I will detail below my specific motivation for carrying out this work on the basis of (a) computational modeling as well as (b) studies of atypically developed individuals, in the specific context of testing the AID hypothesis.

1.5.3.1 *Studies with computational model systems*

In order to systematically examine the impact of early sensory experience on the acquisition of later perceptual proficiencies, one needs to be able to causally interfere with such experience and to examine its consequences rigorously. While this is not feasible in experiments with human participants, considering both ethical and practical reasons, computational models do offer this possibility. As motivated in Section 1.4., deep neural networks represent a particularly attractive class of models. They are generally well-established in the field, are capable of partially accounting for human behavior and neural representations (e.g., Cadena et al., 2019; Khaligh-Razavi and Kriegeskorte, 2014; Storrs et al., 2021), and have been proposed as scientific models and for testing specific neuroscientific hypotheses (e.g., Cichy and Kaiser, 2019; Doerig et al., 2023). Further, unlike many traditional computer vision models operating on the basis of hand-crafted features, a deep neural network is able to learn directly from the inputs it is presented with.

In light of the above, the specific rationale of the modeling approach used in this thesis is as follows: One can train different instances of deep neural networks that only differ in the temporal progression of the sensory inputs to which they are exposed. For instance, in the domain of color vision, as a coarse proxy for biological development, one can train a network by exposing it to color-degraded inputs early in training and to full-color ones later on. One can then compare the trained network to others trained in non-developmental ways (e.g., exclusively on color images, exclusively on color-degraded images, or first on color and later on color-degraded inputs, as an inverse-developmental progression). When comparing the different networks, one can evaluate their proficiency and stability on specific perceptual tasks (e.g., by assessing their ability to recognize objects while modifying color cues) and compare these to known human results. In addition, one can compare the different networks' inner workings through visualization of their learned internal representations, bearing some resemblance to receptive fields in the sensory cortices. Together, such comparisons can help achieve a better understanding of the impact of early sensory experience.

1.5.3.2 *Studies with atypically-developed individuals*

The computational approach above can be complemented by perceptual examinations of special populations of individuals whose early sensory experience deviates from typical development. In the domain of vision, this includes children born blind and treated for their blindness late in life, such as those treated through Project Prakash. Crucially, these individuals differ from normally-sighted children when experiencing the visual world for the first time. This is partly due to many of the maturational processes characteristic of normal development, as reviewed in Section 1.2., proceeding despite the child being blind. Consequently, a late-sighted individual, immediately post-surgically, is equipped with a visual system that is much more mature than that of a newborn, for instance, in terms of acuity (Boas et al., 1969; Hendrickson and Boothe, 1976). Empirical studies of the perceptual abilities of such children, who effectively lack periods of initially degraded vision, can thus help test the functional significance of the early degradations characteristic of typical development. I am fortunate to be able to include some work on Prakash children in the context of testing the AID hypothesis in Chapter 3.

Interestingly, although I could not work with them directly, I was able to link some of the computational findings reported in Chapter 2, focused on testing the AID hypothesis in the auditory domain, to empirical work that has previously been reported in prematurely born babies. Similar to how Prakash individuals lack initial experience with initially degraded inputs, prematurely-born babies have their prenatal periods of exposure to initially low-pass-filtered auditory signals cut short and were immersed in a full-frequency environment earlier than normally-developed newborns (see 2 for details). As such, prematurely born babies represent an additional case of atypical development that is relevant for experimentally validating the AID hypothesis.

1.6 AIM AND STRUCTURE OF THIS THESIS

1.6.1 *Main contributions*

The primary aim of this thesis is to systematically test the AID hypothesis (Section 1.5) based on simulations with deep neural networks and, in part, experiments with atypically developed individuals. I am reporting the results of such examinations for the domains of prenatal hearing (Chapter 2), color vision (Chapter 3), and the joint progression of chromatic sensitivity and visual acuity (Chapter 4), along with a review paper describing and evaluating the AID hypothesis more broadly and in a domain-general fashion (Chapter 5). I provide brief summaries of the aims of these key contributions below.

In Chapter 2, I report computational examinations of the role of initial degradations in the domain of prenatal hearing for the acquisition of later auditory proficiencies. Specifically, as noted earlier, unlike in visual perception, early experience in the auditory modality already begins prenatally. The intrauterine environment thereby effectively acts as a low-pass filter on sounds from the mother's external environment. In this paper, I present computational results of training a deep neural network operating on audio waveforms, using developmentally-inspired and several non-developmental control regimens to examine the consequences of commencing auditory experience with exclusively low-frequency sounds. This work has resulted in the following paper and conference presentation:

- **Vogelsang, M.***, Vogelsang, L.*, Diamond, S., & Sinha, P. (2023). "Prenatal auditory experience and its sequelae". **Published** in Developmental Science, 26(1), e13278. https://doi.org/10.1111/desc.13278. (* = equal contribution)
- **Vogelsang, M.***, Vogelsang, L.*, Diamond, S., & Sinha, P. (2021). "On prenatal auditory experience in humans and its relevance for machine hearing". **Poster presented** at ICLR Workshop "Generalization beyond the training distribution in brains and machines", 2021, Online. (* = equal contribution)

In Chapter 3, I present computational and experimental studies on the role of early visual experience on the later usage of color cues. Newborns begin their visual experience with poor color sensitivity at birth, improving gradually over the following months. To systematically test the consequences of early experience with initially color-degraded inputs on the later usage of color cues, I report results of comprehensive simulations with several deep neural network architectures, databases with different task demands, and different developmentally-inspired and non-developmental training regimens. These computational results are complemented by empirical data on Prakash individuals who were post-surgically tested on their usage of chromatic information. This work has resulted in the following manuscript, which is currently under review:

- **Vogelsang, M.***, Vogelsang, L.*, Gupta, P.*, Gandhi, T., Shah, P., Swami, P., Gilad-Gutnick, S., Ben-Ami, S., Diamond, S., Ganesh, S., & Sinha, P. (Submitted). “Impact of early visual experience on later usage of color cues”. **Under peer-review** at Science (initial submission: September 2023). (* = equal contribution)

In Section 4, I report computational tests of whether the joint progression of visual acuity and color sensitivity early in development helps account for the emergence of the division of the early visual pathway into parvo- and magnocellular systems with distinct response properties. While the latter has long been established as a prominent organizing principle in the mammalian visual system, its genesis is, thus far, unknown. In this paper, I provide a potential account of this genesis based on early sensory development. This account is based on the idea that the joint evolution of features such as spatial frequency and chromatic sensitivities during development may play a causal role in shaping the neural response properties that are characteristic of this division. I provide computational tests to probe this hypothesis. This work resulted in the following submitted manuscript:

- **Vogelsang, M.**, Vogelsang, L., Pipa, G., Diamond, S., & Sinha, P. (Submitted). “On the origin of the parvo- and magnocellular division: potential role of developmental experience”. **Submitted manuscript** (initial submission: October 2023).

In Chapter 5, I present a synthesis of potentially adaptive initial degradations into a domain-general review/perspective paper. While the projects above focus on specific aspects of perceptual development, it is important to review the idea of potentially adaptive initial degradations during development more broadly, based on several of our own studies as well as those of others, the broader historical context, and the potential practical implications for clinical practice and artificial intelligence. Note that, due to an earlier submission timepoint, this paper describes the ideas, but does not yet include the data, of the detailed investigations reported in Chapters 3 and 4. This project has resulted in the following manuscript under review:

- Vogelsang, L.*, **Vogelsang, M.***, Pipa, G., Diamond, S., & Sinha, P. (Submitted). “Butterfly effects in perceptual development: a review of the 'adaptive initial degradation' hypothesis". **Under peer-review** at Developmental Review (initial submission: May 2023). (* = equal contribution)

1.6.2 *Additional contributions reported in the main part of this thesis*

In addition to tests of the AID hypothesis, I have been involved in two additional studies with the Prakash group.

In Chapter 6, I include work on the development of visual memory capacity in Prakash children, to which I contributed through computational modeling with deep neural networks. This work resulted in the following publication:

- Gupta, P., Shah, P., Gilad-Gutnick, S., **Vogelsang, M.**, Vogelsang, L., Tiwari, K., Gandhi, T., Ganesh, S., & Sinha, P. (2022). "Development of visual memory capacity following early-onset and extended blindness". **Published** in Psychological Science, 33(6), 847-858. https://doi.org/10.1177/09567976211056664.

In Chapter 7, I furthermore include examinations of the scholastic performance, and a reflection on the broader educational opportunities, of Prakash children. I contributed non-computationally to this project, together with many co-authors. This work resulted in the following publication:

- Bi, S., Chawariya, A., Ganesh, S., Gupta, P., Huang, Y., Jazayeri, K., Kumar, R., Ralekar, C., Singh, C., Tiwary, A., Vogelsang, L., **Vogelsang, M.**, Yadav, M., & Sinha, P. (2023). "Scholastic status of congenitally blind children following sight surgery". **Published** in International Journal of Special Education, 37(2), 160-168. https://doi.org/10.52291/ijse.2022.37.49. (alphabetical author ordering, except for P. Sinha)

1.6.3 *Additional contributions reported in the appendix of this thesis*

I have also had the chance to computationally contribute to studies of typical development and perceptual processing in normally-sighted adults.

In Appendix A, I include tests of the role of semantics on visual memory capacity in children and adults, to which I contributed through simulations with deep neural networks:

- Gupta, P., Shah, P., Gilad-Gutnick, S., **Vogelsang, M.**, Vogelsang, L., & Sinha, P. (Submitted). "The influence of semantics on visual memory capacity in children and adults". **Under revision** at British Journal of Developmental Psychology (initial submission: December 2022)

In Appendix B, I include a study concerned with the recognition of faces at a distance, to which I primarily contributed through simulations with deep neural networks:

- Jarudi, I. N., Braun, A., **Vogelsang, M.**, Vogelsang, L., Gilad-Gutnick, S., Bosch, X. B., Dixon, III, W. V., & Sinha, P. (2023). "Recognizing distant faces". **Published** in Vision Research, 205, 108184. https://doi.org/10.1016/j.visres.2023.108184.

In Appendix C, I include the conference abstract for the following poster presentation related to the work presented in Chapter 2:

- **Vogelsang, M.***, Vogelsang, L.*, Diamond, S., & Sinha, P. (2021). "On prenatal auditory experience in humans and its relevance for machine hearing". **Poster presented** at ICLR Workshop "Generalization beyond the training distribution in brains and machines", 2021, Online. (* = equal contribution)

REFERENCES

Adams, Russell J and Mary L Courage (2002). "A psychophysical test of the early maturation of infants' mid-and long-wavelength retinal cones." In: *Infant Behavior and Development* 25.2, pp. 247–254.

Allen, Dale, Martin S Banks, and Anthony M Norcia (1993). "Does chromatic sensitivity develop more slowly than luminance sensitivity?" In: *Vision Research* 33.17, pp. 2553–2562.

Allen, Dale, Anthony M Norcia, and Christopher W Tyler (1986). "Comparative study of electrophysiological and psychophysical measurement of the contrast sensitivity function in humans." In: *American journal of optometry and physiological optics* 63.6, pp. 442–449.

Atkinson, J, J Wattam-Bell, E Pimm-Smith, C Evans, and OJ Braddick (1986). "Comparison of rapid procedures in forced choice preferential looking for estimating acuity in infants and young children." In: *Detection and Measurement of Visual Impairment in Pre-Verbal Children: Proceedings of a workshop held at the Institute of Ophthalmology, London on April 1–3, 1985, sponsored by the Commission of the European Communities as advised by the Committed on Medical Research*. Springer, pp. 192–200.

Ayzenberg, Vladislav and Marlene Behrmann (2022). "Development of object recognition." In: *PsyArXiv*.

Azzam, D (2022). "Ronquillo Y. Snellen chart." In: *StatPearls. Treasure island (FL): StatPearls Publishing*.

Bach, Michael et al. (1996). "The Freiburg Visual Acuity Test-automatic measurement of visual acuity." In: *Optometry and vision science* 73.1, pp. 49–53.

Banks, Martin S and Patrick J Bennett (1988). "Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates." In: *JOSA A* 5.12, pp. 2059–2079.

Berkeley, George (1709). *An essay towards a new theory of vision*. Indy-Publish.

Bhalerao, Sushank Ashok, Mahesh Tandon, Satyaprakash Singh, Shraddha Dwivedi, Santosh Kumar, and Jagriti Rana (2015). "Visual impairment and blindness among the students of blind schools in Allahabad and its vicinity: A causal assessment." In: *Indian journal of ophthalmology* 63.3, p. 254.

Bhattacharjee, Harsha, Kalyan Das, Rishi Raj Borah, Kamalesh Guha, Parikshit Gogate, S Purukayastha, and Clare Gilbert (2008). "Causes of childhood blindness in the northeastern states of India." In: *Indian journal of ophthalmology* 56.6, p. 495.

Bieber, Michelle L, Kenneth Knoblauch, and John S Werner (1998). "M-and L-cones in early infancy: II. Action spectra at 8 weeks of age." In: *Vision Research* 38.12, pp. 1765–1773.

Birch, EE, J Gwiazda, JA Bauer Jr, J Naegele, and R Held (1983). "Visual acuity and its meridional variations in children aged 7–60 months." In: *Vision research* 23.10, pp. 1019–1024.

Bishop, Christopher M and Nasser M Nasrabadi (2006). *Pattern recognition and machine learning*. Vol. 4. 4. Springer.

Boas, Judith AR, RL Ramsey, AH Riesen, and JP Walker (1969). "Absence of change in some measures of cortical morphology in dark-reared adult rats." In: *Psychonomic Science* 15.5, pp. 251–252.

Cadena, Santiago A, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, and Alexander S Ecker (2019). "Deep convolutional models improve predictions of macaque V1 responses to natural images." In: *PLoS computational biology* 15.4, e1006897.

Cadieu, Charles F, Ha Hong, Daniel LK Yamins, Nicolas Pinto, Diego Ardila, Ethan A Solomon, Najib J Majaj, and James J DiCarlo (2014). "Deep neural networks rival the representation of primate IT cortex for core visual object recognition." In: *PLoS computational biology* 10.12, e1003963.

Chandna, A (1991). "Natural history of the development of visual acuity in infants." In: *Eye* 5.1, pp. 20–26.

Cichy, Radoslaw M and Daniel Kaiser (2019). "Deep neural networks as scientific models." In: *Trends in cognitive sciences* 23.4, pp. 305–317.

Cichy, Radoslaw Martin, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva (2016). "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence." In: *Scientific reports* 6.1, p. 27755.

Costa, Marcelo Fernandes da, Solange Rios Salomão, Adriana Berezovsky, Filomena Maria de Haro, and Dora Fix Ventura (2004). "Relationship between vision and motor impairment in children with spastic cerebral palsy: new evidence from electrophysiology." In: *Behavioural brain research* 149.2, pp. 145–150.

Crognale, Michael A (2002). "Development, maturation, and aging of chromatic visual pathways: VEP results." In: *Journal of Vision* 2.6, pp. 2–2.

Crognale, Michael A, John P Kelly, AH Weiss, and Davida Y Teller (1998). "Development of the spatio-chromatic visual evoked potential (VEP): a longitudinal study." In: *Vision Research* 38.21, pp. 3283–3292.

Dandona, R and L Dandona (2003). "Childhood blindness in India: a population based perspective." In: *British Journal of Ophthalmology* 87.3, pp. 263–265.

Darwin, Charles (1877). "A biographical sketch of an infant." In: *Mind* 2.7, pp. 285–294.

Daw, Nigel W (2014). *Visual development*. Vol. 14. Springer New York, NY.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei (2009). "Imagenet: A large-scale hierarchical image database." In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Dobkins, Karen R, Barry Lia, and Davida Y Teller (1997). "Infant color vision: Temporal contrast sensitivity functions for chromatic (red/-green) stimuli in 3-month-olds." In: *Vision Research* 37.19, pp. 2699–2716.

Dobson, Velma and Davida Y Teller (1978). "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies." In: *Vision research* 18.11, pp. 1469–1483.

Dobson, Velma, Davida Y Teller, and Jack Belgum (1978). "Visual acuity in human infants assessed with stationary stripes and phase-alternated checkerboards." In: *Vision Research* 18.9, pp. 1233–1238.

Doerig, Adrien, Rowan P Sommers, Katja Seeliger, Blake Richards, Jenann Ismael, Grace W Lindsay, Konrad P Kording, Talia Konkle, Marcel AJ Van Gerven, Nikolaus Kriegeskorte, et al. (2023). "The neuroconnectionist research programme." In: *Nature Reviews Neuroscience*, pp. 1–20.

Dominguez, Melissa and Robert A Jacobs (2003). "Developmental constraints aid the acquisition of binocular disparity sensitivities." In: *Neural Computation* 15.1, pp. 161–182.

Ellemberg, Dave, Terri L Lewis, Chang Hong Liu, and Daphne Maurer (1999). "Development of spatial and temporal vision during childhood." In: *Vision research* 39.14, pp. 2325–2333.

Elman, Jeffrey L (1993). "Learning and development in neural networks: The importance of starting small." In: *Cognition* 48.1, pp. 71–99.

Fantz, Robert L (1958). "Pattern vision in young infants." In: *The psychological record* 8, p. 43.

Fantz, Robert L (1965). "Visual perception from birth as shown by pattern selectivity." In: *Annals of the New York Academy of Sciences*.

Fantz, Robert L, JM Ordy, and MS Udelf (1962). "Maturation of pattern vision in infants during the first six months." In: *Journal of Comparative and Physiological Psychology* 55.6, p. 907.

Fukushima, Kunihiko (1980). "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." In: *Biological cybernetics* 36.4, pp. 193–202.

Gandhi, Tapan K, Amy Kalia Singh, Piyush Swami, Suma Ganesh, and Pawan Sinha (2017). "Emergence of categorical face perception after extended early-onset blindness." In: *Proceedings of the National Academy of Sciences* 114.23, pp. 6139–6143.

Gandhi, Tapan, Amy Kalia, Suma Ganesh, and Pawan Sinha (2015). "Immediate susceptibility to visual illusions after sight onset." In: *Current Biology* 25.9, R358–R359.

Ganesh, Suma, Priyanka Arora, Sumita Sethi, Tapan K Gandhi, Amy Kalia, Garga Chatterjee, and Pawan Sinha (2014). "Results of late surgical intervention in children with early-onset bilateral cataracts." In: *British Journal of Ophthalmology* 98.10, pp. 1424–1428.

Gardner, R and ED Weitzman (1967). "Examination for optokinetic nystagmus in sleep and waking." In: *Archives of Neurology* 16.4, pp. 415–420.

Geirhos, Robert, Kantharaju Narayanappa, Benjamin Mitzkus, Tizian Thieringer, Matthias Bethge, Felix A Wichmann, and Wieland Brendel (2021). "Partial success in closing the gap between human and machine vision." In: *Advances in Neural Information Processing Systems* 34, pp. 23885–23899.

Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel (2018a). "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness." In: *arXiv preprint arXiv:1811.12231*.

Geirhos, Robert, Carlos RM Temme, Jonas Rauber, Heiko H Schütt, Matthias Bethge, and Felix A Wichmann (2018b). "Generalisation in humans and deep neural networks." In: *Advances in neural information processing systems* 31.

Geldart, Sybil, Catherine J Mondloch, Daphne Maurer, Scania De Schonen, and Henry P Brent (2002). "The effect of early visual deprivation on the development of face processing." In: *Developmental Science* 5.4, pp. 490–501.

Gerhardt, Kenneth J and Robert M Abrams (1996). "Fetal hearing: characterization of the stimulus and response." In: *Seminars in perinatology*. Vol. 20. 1. Elsevier, pp. 11–20.

Gibson, James J (1979). *The ecological approach to visual perception: classic edition*. Psychology press.

Gibson, James Jerome (1966). *The senses considered as perceptual systems.* Houghton Mifflin.

Gilbert, CE, R Canovas, M Hagan, S Rao, and A Foster (1993). "Causes of childhood blindness: results from west Africa, south India and Chile." In: *Eye* 7.1, pp. 184–188.

Gilbert, Clare and Haroon Awan (2003). *Blindness in children*.

Gilbert, Clare and Allen Foster (2001). "Childhood blindness in the context of VISION 2020: the right to sight." In: *Bulletin of the World Health Organization* 79.3, pp. 227–232.

Good, William V (2001). "Development of a quantitative method to measure vision in children with chronic cortical visual impairment." In: *Transactions of the American Ophthalmological Society* 99, p. 253.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press.

Granrud, Carl E, Robert J Haake, and Albert Yonas (1985). "Infants' sensitivity to familiar size: The effect of memory on spatial perception." In: *Perception & psychophysics* 37.5, pp. 459–466.

Griffiths, Scott K, WS Brown Jr, Kenneth J Gerhardt, Robert M Abrams, and Richard J Morris (1994). "The perception of speech sounds recorded within the uterus of a pregnant sheep." In: *The Journal of the Acoustical Society of America* 96.4, pp. 2055–2063.

Haber, Ralph Norman (1985). "Perception: A one-hundred-year perspective." In: *A century of psychology as science*, pp. 224–230.

Hamer, Russell D, Anthony M Norcia, Christopher W Tyler, and Charlene Hsu-Winges (1989). "The development of monocular and binocular VEP acuity." In: *Vision Research* 29.4, pp. 397–408.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016). "Deep residual learning for image recognition." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hebb, Donald O (1949). "The first stage of perception: growth of the assembly." In: *The Organization of Behavior* 4.60, pp. 78–60.

Heersema, DJ and J Vanhofvanduin (1990). "Age norms for visual-acuity in toddlers using the acuity card procedure." In: *Clinical Vision Sciences* 5.2, pp. 167–174.

Held, Richard, Yuri Ostrovsky, Beatrice de Gelder, Tapan Gandhi, Suma Ganesh, Umang Mathur, and Pawan Sinha (2011). "The newly sighted fail to match seen with felt." In: *Nature neuroscience* 14.5, pp. 551–553.

Hendrickson, Anita and Ronald Boothe (1976). "Morphology of the retina and dorsal lateral geniculate nucleus in dark-reared monkeys (Macaca nemestrina)." In: *Vision research* 16.5, 517–IN5.

Hepper, Peter G and B Sara Shahidullah (1994). "The development of fetal hearing." In: *Fetal and Maternal Medicine Review* 6.3, pp. 167–179.

Hering, Ewald (1861). *Beitrage zur physiologie*. W. Engelmann.

Hinton, Geoffrey E, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov (2012). "Improving neural networks by preventing co-adaptation of feature detectors." In: *arXiv preprint arXiv:1207.0580*.

Hubel, David H and Torsten N Wiesel (1959). "Receptive fields of single neurones in the cat's striate cortex." In: *The Journal of physiology* 148.3, p. 574.

Hubel, David H and Torsten N Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." In: *The Journal of physiology* 160.1, p. 106.

Hubel, David Hunter, Torsten Nils Wiesel, Simon LeVay, Horace Basil Barlow, and Raymond Michael Gaze (1977). "Plasticity of ocular dominance columns in monkey striate cortex." In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 278.961, pp. 377–409.

Ioffe, Sergey and Christian Szegedy (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In: *International conference on machine learning*. pmlr, pp. 448–456.

James, William (1890). *The principles of psychology*. Vol. 1. Cosimo, Inc.

Kalia, Amy, Tapan Gandhi, Garga Chatterjee, Piyush Swami, Harvendra Dhillon, Shakeela Bi, Naval Chauhan, Shantanu Das Gupta, Preeti Sharma, Saahil Sood, et al. (2017). "Assessing the impact of a program for late surgical intervention in early-blind children." In: *Public Health* 146, pp. 15–23.

Kalia, Amy, Luis Andres Lesmes, Michael Dorr, Tapan Gandhi, Garga Chatterjee, Suma Ganesh, Peter J Bex, and Pawan Sinha (2014). "Development of pattern vision following early and extended blindness." In: *Proceedings of the National Academy of Sciences* 111.5, pp. 2035–2039.

Kellman, Philip J and Martha E Arterberry (2007). "Infant visual perception." In: *Handbook of child psychology* 2.

Kelly, John P, Katja Borchert, and Davida Y Teller (1997). "The development of chromatic and achromatic contrast sensitivity in infancy as tested with the sweep VEP." In: *Vision Research* 37.15, pp. 2057–2072.

Khaligh-Razavi, Seyed-Mahdi and Nikolaus Kriegeskorte (2014). "Deep supervised, but not unsupervised, models may explain IT cortical representation." In: *PLoS computational biology* 10.11, e1003915.

Khanna, Rohit C (2018). "Commentary: Childhood blindness in India: Regional variations." In: *Indian Journal of Ophthalmology* 66.10, p. 1461.

Kietzmann, Tim C, Courtney J Spoerer, Lynn KA Sörensen, Radoslaw M Cichy, Olaf Hauk, and Nikolaus Kriegeskorte (2019). "Recurrence is required to capture the representational dynamics of the human visual system." In: *Proceedings of the National Academy of Sciences* 116.43, pp. 21854–21863.

Kiorpes, Lynne (2015). "Visual development in primates: neural mechanisms and critical periods." In: *Developmental neurobiology* 75.10, pp. 1080–1090.

Kiorpes, Lynne (2016). "The puzzle of visual development: behavior and neural limits." In: *Journal of Neuroscience* 36.45, pp. 11384–11393.

Kiorpes, Lynne and J Anthony Movshon (2004). "Neural limitations on visual development in primates." In: *The visual neurosciences* 1, pp. 159–173.

Knoblauch, Kenneth, Michelle L Bieber, and John S Werner (1998). "M- and L-cones in early infancy: I. VEP responses to receptor-isolating stimuli at 4-and 8-weeks of age." In: *Vision Research* 38.12, pp. 1753–1764.

Kriegeskorte, Nikolaus, Marieke Mur, and Peter A Bandettini (2008). "Representational similarity analysis-connecting the branches of systems neuroscience." In: *Frontiers in systems neuroscience*, p. 4.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems* 25.

Kubilius, Jonas, Stefania Bracci, and Hans P Op de Beeck (2016). "Deep neural networks as a computational model for human shape sensitivity." In: *PLoS computational biology* 12.4, e1004896.

Le Vay, Simon, Torsten N Wiesel, and David H Hubel (1980). "The development of ocular dominance columns in normal and visually deprived monkeys." In: *Journal of Comparative Neurology* 191.1, pp. 1–51.

LeCun, Yann, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel (1989). "Backpropagation applied to handwritten zip code recognition." In: *Neural computation* 1.4, pp. 541–551.

Leat, Susan J, Naveen K Yadav, and Elizabeth L Irving (2009). "Development of visual acuity and contrast sensitivity in children." In: *Journal of optometry* 2.1, pp. 19–26.

Locke, John (1690). "An essay concerning human understanding." In: *Readings in the history of psychology*. Appleton-Century-Crofts, pp. 55–68.

Mandavilli, Apoorva (2006). "Visual neuroscience: look and learn." In: *Nature* 441.7091, pp. 271–273.

Marg, E, DN Freeman, and PJ Goldstein (1976). "Visual acuity development in human infants: evoked potential measurements." In: *Invest Ophthalmol* 15, pp. 150–153.

Mayer, D Luisa and Velma Dobson (1982). "Visual acuity development in infants and young children, as assessed by operant preferential looking." In: *Vision research* 22.9, pp. 1141–1151.

McCulloch, Warren S and Walter Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." In: *The bulletin of mathematical biophysics* 5, pp. 115–133.

McDonald, MA, V Dobson, SL Sebris, L Baitch, D Varner, and DY Teller (1985). "The acuity card procedure: a rapid test of infant acuity." In: *Investigative ophthalmology & visual science* 26.8, pp. 1158–1162.

McDonald, Maryalice, Lawson S Sebris, Gesine Mohn, Davida Y Teller, and Velma Dobson (1986). "Monocular acuity in normal infants: the acuity card procedure." In: *Optometry and Vision Science* 63.2, pp. 127–134.

Mehrer, Johannes, Courtney J Spoerer, Emer C Jones, Nikolaus Kriegeskorte, and Tim C Kietzmann (2021). "An ecologically motivated image dataset for deep learning yields better models of human vision." In: *Proceedings of the National Academy of Sciences* 118.8, e2011417118.

Mill, John Stuart (1865). "An Examination of Sir William Hamilton's Philosophy (1865)." In: *Toronto, University of Toronto*, p. 184.

Minsky, ML and SA Papert (1969). *Perceptrons. An Introduction to Computational Geometry. 1969, Expanded.*

Murkoff, Heidi (2016). *What to expect when you're expecting*. Workman Publishing.

Neu, Beate and Ruxandra Sireteanu (1997). "Monocular acuity in preschool children: assessment with the Teller and Keeler acuity cards in comparison to the C-test." In: *Strabismus* 5.4, pp. 185–202.

Newport, Elissa L (1988). "Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language." In: *Language sciences* 10.1, pp. 147–172.

Norcia, Anthony M, L Gregory Appelbaum, Justin M Ales, Benoit R Cottereau, and Bruno Rossion (2015). "The steady-state visual evoked potential in vision research: A review." In: *Journal of vision* 15.6, pp. 4–4.

Norcia, Anthony M and Christopher W Tyler (1985). "Spatial frequency sweep VEP: visual acuity during the first year of life." In: *Vision research* 25.10, pp. 1399–1408.

Norcia, Anthony M, Christopher W Tyler, and Dale Allen (1985). "Electrophysiological assessment of contrast sensitivity in human infants." In: *Noninvasive Assessment of Visual Function*. Optica Publishing Group, WB2.

Norman, Joel (2002). "Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches." In: *Behavioral and brain sciences* 25.1, pp. 73–96.

Ostrovsky, Yuri, Aaron Andalman, and Pawan Sinha (2006). "Vision following extended congenital blindness." In: *Psychological Science* 17.12, pp. 1009–1014.

Ostrovsky, Yuri, Ethan Meyers, Suma Ganesh, Umang Mathur, and Pawan Sinha (2009). "Visual parsing after recovery from blindness." In: *Psychological Science* 20.12, pp. 1484–1491.

Peeles, David R and Davida Y Teller (1975). "Color vision and brightness discrimination in two-month-old human infants." In: *Science* 189.4208, pp. 1102–1103.

Piller, Sophia, Irene Senna, Dennis Wiebusch, Itay Ben-Zion, and Marc O Ernst (2023). "Grasping behavior does not recover after sight restoration from congenital blindness." In: *Current Biology* 33.10, pp. 2104–2110.

Rahi, JS, CE Gilbert, A Foster, and D Minassian (1999). "Measuring the burden of childhood blindness." In: *British journal of ophthalmology* 83.4, pp. 387–388.

Regan, D (1973). "Rapid objective refraction using evoked brain potentials." In: *Investigative Ophthalmology & Visual Science* 12.9, pp. 669–679.

Rosenblatt, Frank (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1986). "Learning representations by back-propagating errors." In: *nature* 323.6088, pp. 533–536.

Schrimpf, Martin, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo (2020a). "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?" In: *bioRxiv*.

Schrimpf, Martin, Jonas Kubilius, Michael J Lee, N Apurva Ratan Murty, Robert Ajemian, and James J DiCarlo (2020b). "Integrative benchmarking to advance neurally mechanistic models of human intelligence." In: *Neuron* 108.3, pp. 413–423.

Senna, Irene, Elena Andres, Ayelet McKyton, Itay Ben-Zion, Ehud Zohary, and Marc O Ernst (2021). "Development of multisensory integration following prolonged early-onset visual deprivation." In: *Current Biology* 31.21, pp. 4879–4885.

Sinha, Pawan (2013). "Once blind and now they see." In: *Scientific American* 309.1, pp. 48–55.

Sinha, Pawan (2016). "Neuroscience and service." In: *Neuron* 92.3, pp. 647–652.

Sinha, Pawan, Garga Chatterjee, Tapan Gandhi, and Amy Kalia (2013). "Restoring vision through "Project Prakash": the opportunities for merging science and service." In: *PLoS Biology* 11.12, e1001741.

Sinha, Pawan and Richard Held (2012). "Sight restoration." In: *F1000 Medicine Reports* 4.

Sireteanu, Ruxandra (2000). "Development of the visual system in the human infant." In: *Handbook of brain and behaviour in human development*. Ed. by AF Kalverboer and A Gramsberger. Kluwer Academic Publishers Dordrecht, Boston, London, pp. 629–652.

Smith, Sherri L, Kenneth J Gerhardt, Scott K Griffiths, Xinyan Huang, and Robert M Abrams (2003). "Intelligibility of sentences recorded from the uterus of a pregnant ewe and from the fetal inner ear." In: *Audiology and Neurotology* 8.6, pp. 347–353.

Sokol, Samuel (1976). "Visually evoked potentials: theory, techniques and clinical applications." In: *Survey of ophthalmology* 21.1, pp. 18–44.

Spoerer, Courtney J, Patrick McClure, and Nikolaus Kriegeskorte (2017). "Recurrent convolutional neural networks: a better model of biological object recognition." In: *Frontiers in psychology* 8, p. 1551.

Steinkuller, Paul G, Lee Du, Clare Gilbert, Allen Foster, Mary Louise Collins, and David K Coats (1999). "Childhood blindness." In: *Journal of American Association for Pediatric Ophthalmology and Strabismus* 3.1, pp. 26–32.

Stiers, Peter, Ria Vanderkelen, and Erik Vandenbussche (2003). "Optotype and grating visual acuity in preschool children." In: *Investigative ophthalmology & visual science* 44.9, pp. 4123–4130.

Storrs, Katherine R, Tim C Kietzmann, Alexander Walther, Johannes Mehrer, and Nikolaus Kriegeskorte (2021). "Diverse deep neural networks all predict human inferior temporal cortex well, after training and fitting." In: *Journal of cognitive neuroscience* 33.10, pp. 2044–2064.

Suttle, Catherine M, Martin S Banks, and Erich W Graf (2002). "FPL and sweep VEP to tritan stimuli in young human infants." In: *Vision Research* 42.26, pp. 2879–2891.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015). "Going deeper with convolutions." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus (2013). "Intriguing properties of neural networks." In: *arXiv preprint arXiv:1312.6199*.

Tang, Hanlin, Martin Schrimpf, William Lotter, Charlotte Moerman, Ana Paredes, Josue Ortega Caro, Walter Hardesty, David Cox, and Gabriel Kreiman (2018). "Recurrent computations for visual pattern completion." In: *Proceedings of the National Academy of Sciences* 115.35, pp. 8835–8840.

Teller, Davida Y (1979). "The forced-choice preferential looking procedure: A psychophysical technique for use with human infants." In: *Infant Behavior and Development* 2, pp. 135–153.

Teller, Davida Y (1998). "Spatial and temporal aspects of infant color vision." In: *Vision Research* 38.21, pp. 3275–3282.

Teller, Davida Y, Ralph Morse, Richard Borton, and David Regal (1974). "Visual acuity for vertical and diagonal gratings in human infants." In: *Vision research* 14.12, pp. 1433–1439.

Turkewitz, Gerald and Patricia A Kenny (1982). "Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement." In: *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 15.4, pp. 357–368.

Tyler, Christopher W, Patricia Apkarian, Dennis M Levi, and Ken Nakayama (1979). "Rapid assessment of visual function: an electronic sweep technique for the pattern visual evoked potential." In: *Investigative Ophthalmology & Visual Science* 18.7, pp. 703–713.

Ver Hoeve, James N, Thomas D France, and G Andrew Bousch (1996). *A sweep VEP test for color vision deficits in infants and young children.*

Vogelsang, Lukas, Sharon Gilad-Gutnick, Evan Ehrenberg, Albert Yonas, Sidney Diamond, Richard Held, and Pawan Sinha (2018). "Potential downside of high initial visual acuity." In: *Proceedings of the National Academy of Sciences* 115.44, pp. 11333–11338.

Wiesel, Torsten N and David H Hubel (1965). "Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens." In: *Journal of neurophysiology* 28.6, pp. 1029–1040.

World Health Organization (2000). *Preventing blindness in children: report of a WHO/IAPB scientific meeting.*

Ye, Jie, Priti Gupta, Pragya Shah, Kashish Tiwari, Tapan Gandhi, Suma Ganesh, Flip Phillips, Dennis Levi, Frank Thorn, Sidney Diamond, et al. (2021). "Resilience of temporal processing to early and extended visual deprivation." In: *Vision Research* 186, pp. 80–86.

Yonas, Albert, Linda Pettersen, and Carl E Granrud (1982). "Infants' sensitivity to familiar size as information for distance." In: *Child Development*, pp. 1285–1290.

Yuodelis, Cristine and Anita Hendrickson (1986). "A qualitative and quantitative analysis of the human fovea during development." In: *Vision research* 26.6, pp. 847–855.

Zheng, Xiaowei, Guanghua Xu, Kai Zhang, Renghao Liang, Wenqiang Yan, Peiyuan Tian, Yaguang Jia, Sicong Zhang, and Chenghang Du (2020). "Assessment of human visual acuity using visual evoked potential: A review." In: *Sensors* 20.19, p. 5542.

Part II

RESEARCH

2 PRENATAL AUDITORY EXPERIENCE AND ITS SEQUELAE

Content from

Vogelsang, M.*, Vogelsang, L.*, Diamond, S., & Sinha, P. (2023). "Prenatal auditory experience and its sequelae". **Published** in Developmental Science, 26(1), e13278. (* = equal contribution)

2.1 ABSTRACT

Towards the end of the second trimester of gestation, a human fetus is able to register environmental sounds. This in utero auditory experience is characterized by comprising strongly low-pass-filtered versions of sounds from the external world. Here, we present computational tests of the hypothesis that this early exposure to severely degraded auditory inputs serves an adaptive purpose—it may induce the neural development of extended temporal integration. Such integration can facilitate the detection of information carried by low-frequency variations in the auditory signal, including emotional or other prosodic content. To test this prediction, we characterized the impact of several training regimens, biomimetic and otherwise, on a computational model system trained and tested on the task of emotion recognition. We find that training with an auditory trajectory recapitulating that of a neurotypical infant in the pre-to-postnatal period results in temporally extended receptive field structures and yields the best subsequent accuracy and generalization performance on the task of emotion recognition. This strongly suggests that the progression from low-passfiltered to full-frequency inputs is likely to be an adaptive feature of our development, conferring significant benefits to later auditory processing abilities relying on temporally extended analyses. Additionally, this finding can help explain some of the auditory impairments associated with preterm births, suggests guidelines for the design of auditory environments in neonatal care units, and points to enhanced training procedures for computational models.

2.2 INTRODUCTION

Expectant mothers often report that their unborn child exhibits movements in response to loud environmental sounds (Murkoff, 2016). Further attesting to these anecdotal reports, systematic developmental psychoacoustic studies have demonstrated that by around 20 weeks of gestational age, the fetus is equipped with a functioning auditory system, capable of registering the mother's soundscape (Hepper and Shahidullah, 1994) (see Figure 2.1a), and may even benefit from such auditory exposure, especially to the mother's vocalizations (Webb et al., 2015). However, the quality of this auditory experience is quite limited. Factors such as the fluid medium and surrounding tissues in the intrauterine environment, the impedance of the fetal skull, and the immaturity of the cochlea lead to a strong reduction of the audibility of high frequencies present in environmental sounds while affecting lower frequencies only marginally or even enhancing them (Gerhardt and Abrams, 1996; Griffiths et al., 1994; Hepper and Shahidullah, 1994). In other words, the intrauterine environment essentially acts as a low-pass filter on incident sounds, significantly limiting the fetus' exposure and sensitivity to high-frequency auditory stimuli. This is further illustrated in Figure 2.1a.

The question we consider in this paper is whether these degradations of incident sounds are epiphenomenal limitations imposed by biological processes or if they may, in fact, have any salutary implications later in life. In other words, might the degradation in prenatal experience not limit a child's later auditory skills but possibly enhance them? This question has the potential to contribute to a broader and mostly unresolved debate in the field of developmental science, concerned with whether developmental constraints are to be considered as hindering, or rather helping, the acquisition of later perceptual abilities. In the traditional view, early limitations are typically seen as the outcome of constraints imposed by the physiological immaturities of the underlying neural and perceptual systems, and human development is believed to somehow have to surmount the challenges imposed by these early limitations to attain its later manifesting proficiencies. This perspective, despite its perhaps intuitive appeal, has been challenged in specific domains of investigation. The proposal put forward by Turkewitz and Kenny (1982) was among the first to suggest that early limitations, rather than solely representing hurdles, may, in fact, be adaptive for later developmental proficiencies. Specifically, Turkewitz and Kenny (1982) proposed that by initially reducing the complexity of environmental stimuli, a learner's perceptual analysis would be rendered less overwhelming. Along similar lines, but focused specifically on learning, and using computational simulations to support their claims, other researchers have suggested, for instance, that developmental limitations in cognitive architectures can benefit language learning (Elman, 1993; Newport, 1988) and that early limitations in an infant's visual system can benefit the acquisition of binocular disparity detection capabilities (Dominguez and Jacobs, 2003). Here, we wish to extend this ongoing investigation to the domain of auditory processing as well as to prenatal, rather than exclusively postnatal, development.

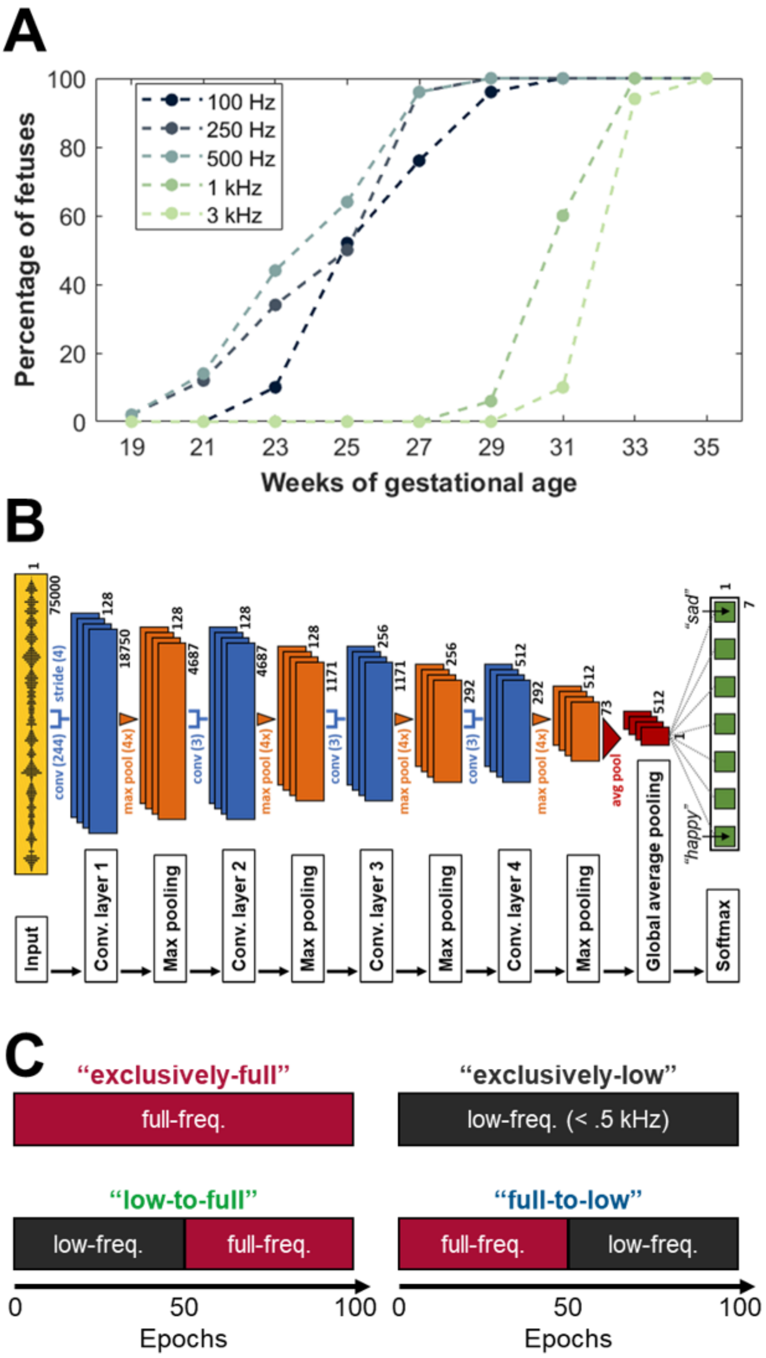


Figure 2.1: (a) Prenatal auditory sensitivity: percentage of fetuses that responded to each of the given frequencies at different gestational ages, adapted from Hepper and Shahidullah (1994). (b) Architecture of the network "M5" by Dai et al. (2017), equipped with four convolutional layers, followed by global average pooling and a classification layer containing seven output nodes of the network, representing the seven different emotion classes to be predicted (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral emotion). (c) Schematics of the four training regimens used, each comprising 100 epochs of training

In the domain of auditory processing, the following logic makes a case for a way in which early degradations may be adaptive: By definition, low-pass-filtered audio streams, roughly mirroring the fetal experience in the uterus, contain markedly reduced fine-grained, high frequency information. As a consequence of this reduction of higher frequency content that is available to the auditory system at any moment in time, brief audio snippets are rendered relatively uninformative for inference upon information and events in the environment. In order to, nevertheless, derive meaningful inferences, a strategy that the system may learn to employ is to gather information across longer time scales, which would consequently induce the development of neural mechanisms capable of integrating information over extended time spans. Such temporal integration is known to, later in life, be useful for the processing of information carried by slow variations in the auditory signal (i.e., those relying on temporally extended analysis), including the expression of emotions or other prosodic content (Ross et al., 1973; Snel and Cullen, 2013). Contrasting this perspective, the early availability of high frequencies in the auditory environment may render even short audio snippets sufficiently informative for the perceptual system and may thus preclude the development of extended temporal integration and, hence, compromise specific auditory skills that depend on it. Taken together, this leads us to hypothesize that degraded, approximately low-pass-filtered, inputs play an adaptive role in (i) instantiating extended temporal integration mechanisms and, (ii) as a consequence thereof, enable robust auditory analysis for tasks relying on temporally extended information, such as the recognition of emotions or other prosodic content.

To systematically probe these two predictions, and thereby evaluate our proposal's overall plausibility, we need to be able to actively manipulate experience and assess the consequences of these manipulations. While ethical and practical considerations render such deliberate manipulation infeasible in human participants, studies with computational model systems, in particular deep convolutional neural networks (DCNNs), which are capable of directly learning from experience, offer a powerful way forward. These networks, while not perfect models of the biological system, can serve as useful approximations of early sensory processing (Norman-Haignere and McDermott, 2018) and allow for a systematic assessment of the consequences of experience with different types of stimuli. Specifically, we can expose a deep convolutional network to different temporal progressions of auditory inputs (some of which are designed to recapitulate biological development, while others are not) and examine the resulting differences in outcomes.

The following seeks to illustrate how we can probe the two key predictions introduced above through an examination of these outcomes: In brief, a network such as the one we are using (see Figure 2.1b for an illustration) takes short audio recordings of spoken utterances as inputs and, in rough analogy to the biological system, processes these through a hierarchical cascade of temporal filters. Due to this organizational scheme, as one proceeds along the processing stages, more and more complex speech properties are extracted. These properties, in turn, are learned to be associated with classification labels (in our case, the emotion that a given utterance was spoken with). These associations are initially random but are progressively refined during the training of the network: As part of this training procedure, numerous audio clips are repeatedly fed into the network along with the desired classification label (e.g., "sad" or "happy"), and the connections are refined so as to reduce classification error. Crucially, not only the links between extracted higher-level features and the final classification labels, but also the temporal filters in the early processing layers themselves, which are in rough analogy to temporal receptive fields found in the auditory cortex, are learned. Following the same principle, initially random filters are sculpted throughout the training process, so as to be able to extract informative features of the auditory signal. Following training with stimulus progressions that are either recapitulating biological development or not (see "2.3.3 Training Regimens" for details), the resulting receptive field structures can be compared, allowing us to probe the first key prediction we had stated earlier: that training with degraded, approximately low-pass-filtered, inputs helps instantiate extended temporal integration mechanisms (i.e., lead to receptive fields encoding lower frequencies and thereby encoding longer wavelengths). In addition, we can also probe our second key prediction, that these kinds of inputs, as a consequence of instantiating extended temporal integration mechanisms, enable robust auditory analysis for tasks relying on temporally extended information. This can directly be examined based on the networks' performances and generalization behavior on the task of emotion classification—a task that bears great ecological significance for humans and also relies on temporally extended auditory analysis.

To sum, the main goal of our computational investigation is to expose different instances of our network to different temporal progressions of filtered speech signals (one of which is inspired by the developmental progression from low-pass-filtered to full-frequency inputs, and three other progressions serving as nondevelopmental controls; see "2.3.3 Training Regimens" for details) while evaluating the two above-mentioned aspects of the resulting networks: (i) their temporal integration profiles (as is evident from the resulting receptive field structures), and (ii) their performances and generalization abilities on a task relying on temporally extended analysis (specifically, emotion classification).

2.3 methods

2.3.1 *Computational model*

As motivated in the Introduction, we utilized a DCNN as our computational model system for this study, and trained it on classifying short audio clips into one of seven emotion categories. More specifically, we used the model "M5" by Dai et al. (2017), whose architecture is sketched in Figure 2.1b. This model allows feeding audio recordings of spoken utterances, represented as raw audio waveforms, as inputs. These inputs are subsequently processed through a cascade of temporal filters. This is implemented through a sequence of four convolutional layers—the first of which takes the raw audio clip as input, the later ones taking the outputs of the previous layer as inputs. As part of each convolutional layer, the layer's input is convolved with a set of temporal filters, resulting in a set of outputs, which are subsequently down-sampled through the application of the max pooling operation. Due to this hierarchical processing scheme, as the audio signal is propagating through the convolutional layers of the network, more and more complex auditory features can be extracted. Finally, the outputs of the last convolutional layer, after passing through an additional pooling stage, are connected to the output nodes of the final classification layer, whose activities represent the probabilities for a given audio clip to belong to each of the seven different emotion categories that the network was trained on differentiating.

The shape of the convolutional filters as well as the connectivity patterns in the final classification layer are initialized randomly but get sculpted during training, as part of which the audio inputs and their desired classification labels are fed into the network, and the network weights (comprising the filters and final connectivity patterns) are progressively adjusted, so as to reduce the classification error. Following this training phase, the learned filters, as well as the resulting performances on the emotion classification task, can be examined.

2.3.2 *Dataset for training and testing our model*

As dataset for training and testing our network, we utilized the Toronto Emotional Speech Set (Dupuis and Pichora-Fuller, 2010). This database contains 200 spoken utterances for each of seven different emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral emotion) and two different speakers. The individual audio snippets span durations between 1 and 3 s, and recordings were accordingly appended with silence in order to be of identical length.

The dataset, containing 2800 audio clips in total, was split into a "training set" containing 90% of the audio clips and a "test set" containing the remaining 10% of the clips. As the names indicate, the training set was used to train the network (i.e., to have the network adjust its weights while repeatedly feeding it the audio clips from the training set, together with the desired emotion classification labels), and the test set used to evaluate the ability of the network to correctly classify audio clips that the network was not explicitly exposed to during training.

To investigate the stability of our findings related to performance, we trained and tested each network not only once but for a total of 10 times, applying the following systematicity: The database was split to produce 10 versions of training sets each containing 90% of the clips, as well as test sets each containing the other 10%, while ensuring that the audio clips in the different test sets were not overlapping. The networks were then trained and tested on all 10 variations (which we also refer to as "folds"), and the performance-based Figures 2.4a,b was plotted to depict the mean and standard error when averaging across these. Note that Figures 2.3b,2.3c,and 2.4c represent the results obtained by pooling together the outcomes of all 10 folds, and that Figures 2.2a–2.2c and 2.3a depict results of a single fold (the first one), so as to allow the concrete visualization of individual filters' data points, instead of providing summary statistics across folds.

2.3.3 *Training regimens*

To test our two key predictions, we trained different instances of our network on the task of classifying a given audio recording into one of seven emotions, based on the Toronto Emotional Speech Set (Dupuis and Pichora-Fuller, 2010), using four qualitatively different regimens. One of these regimens was designed to recapitulate biological development in its temporal progression, while three others served as non-developmental controls. The four different regimens, as illustrated in Figure 2.1c, are:

- "Low-to-full", in which, inspired by the prenatal-to-postnatal progression in development, training commenced with low-pass-filtered and resumed with full-frequency inputs,
- "Full-to-low", in which, as a control condition, training followed an inverse-developmental progression but had the same aggregate content as the previous ("low-tofull") regimen,
- "Exclusively-full", in which, more simply, the entire training consisted of full-frequency input, as an additional control, and
- "Exclusively-low", in which, similarly simply, the entire training was based on low-pass-filtered input, as a final control.

Each training regimen thereby comprised a total of 100 training epochs (representing the number of times that the entire input data was fed into the network during training), with the two different stages in the "low-to-full" and "full-to-low" regimens consisting of 50 epochs each. Low-pass filtering was carried out at a cutoff frequency of 500 Hz, as inspired by previous recordings in the womb reported in Hepper and Shahidullah (1994) and illustrated in Figure 2.1a. The specifics of training and analysis procedures are detailed in the Supporting Information.

2.4 RESULTS

To assess the impact of training regimen on our networks' early, learned representations, we examined the spectral profiles of their temporal receptive fields in the first convolutional layer. This analysis revealed that while the network trained on full-frequency inputs learned receptive fields with a broad range of peak frequencies, the network trained on low-pass-filtered data acquired exclusively low-frequency receptive field structures (see Figure 2.2a,b). Further illustrating this point, Figure 2.2c shows spectral profiles of specific individual filters (those with sorting indices corresponding to the x-axis ticks in Figure 2.2a,b) for training following exclusively-low and exclusively-full regimens.

Further, training on low-frequency inputs resulted not only in more, but also purer, low-frequency filters. For filters with peak frequencies up to 0.5 kHz (126 filters following exclusively-low training, and 83 filters following exclusively-full training), there was a significant difference across the two training conditions in the contribution of frequencies up to 0.5 kHz to the filter responses ($t(207) = 13.91$, $p < 0.001$ for two-tailed t-test; see Figure 2.3a). This correspondence, as evident in Figures 2.2 and 2.3a, attests to the idea that initial exposure to exclusively low-frequency content, as is characteristic of prenatal hearing, may enforce the development of extended temporal integration mechanisms.

Moving beyond homogenous training regimens, we next studied the impact of biomimetic and reverse-biomimetic protocols on the network's learned representations in the first convolutional layer. Specifically, we examined changes in receptive field properties as the network transitioned from one phase of training (low- or full-frequency inputs) to the other (full- or low-frequency inputs). The results revealed that while only 13% of the receptive fields established during the first half of training on low-frequency inputs changed their peak frequencies upon transitioning to full-frequency inputs, this was the case for 45% of the receptive field structures when full-frequency training was later followed by low-frequency training (see Figure 2.3b). That these filters were almost exclusively enlarged temporally, further resulted in the full-to-low model approaching the spectral distributions of the exclusively-low and low-to-full models (see Figure 2.3c). These results

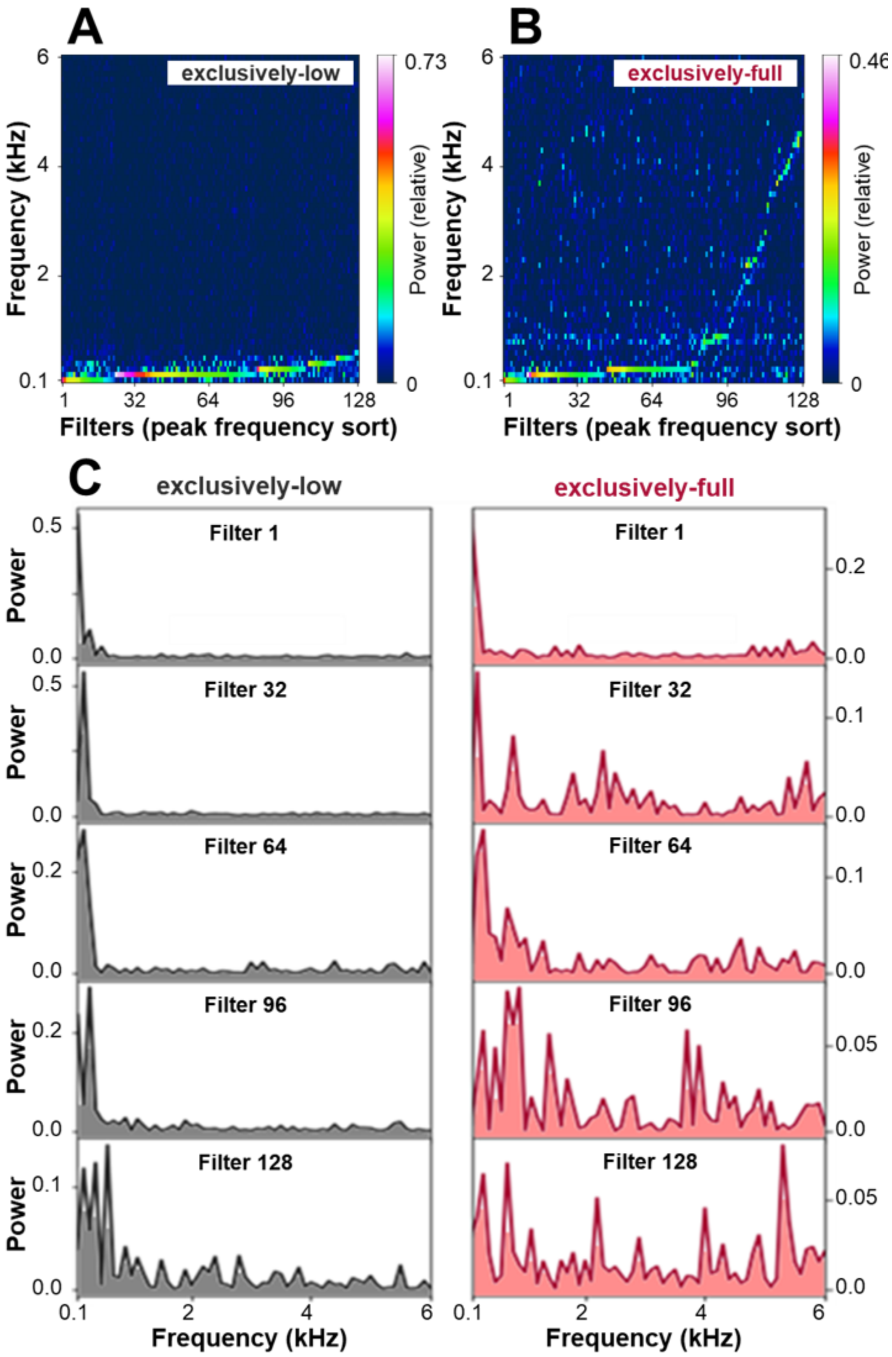


Figure 2.2: Receptive field analysis. (a) and (b) Spectral distribution of first-layer filters in the network trained on full-frequency (a) and low-frequency (b) input. Colors code for normalized power obtained through frequency decomposition, with the sum of values up to 6 kHz normalized to 1. (c) Close-up of spectral profiles of individual filters (filters with the 1st, 32nd, 64th, 96th, and 128th lowest peak frequency), when training followed exclusively-low (left) and exclusively-full (right) regimens, respectively

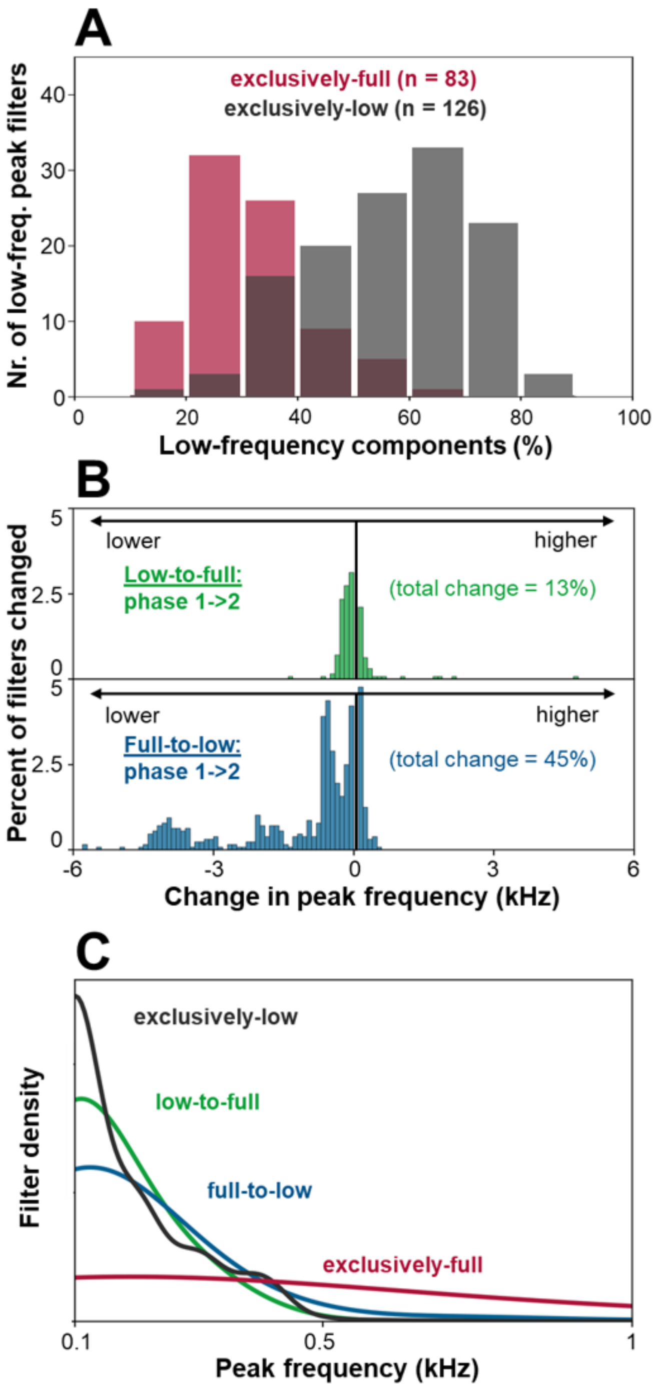


Figure 2.3: Additional receptive field analysis. (a) Histogram of the proportion that frequencies up to 0.5 kHz contribute, relative to all frequencies up to 6 kHz, to the response of individual filters with low peak frequencies (up to 0.5 kHz), following exclusively-low and exclusively-full training. (b) Histogram of units whose peak frequency changed from the first to the second half of training, separately for low-to-full and full-to-low training (pooled across 10-fold variations; see Methods section). (c) Kernel density estimation plot of the distribution of filters' peak frequencies for all networks (pooled across 10-fold variations)

suggest that temporally extended receptive fields, acquired through initial low-pass-filtered training, may comprise more robustly useful processing units for the task, rendering the need for their later adjustment superfluous.

Complementing these results, we next examined the consequences of our four training regimens on the networks' later generalization performance by testing emotion classification on full-frequency and various low-pass-filtered test data. The results depicted in Figure 2.4a reveal that while the model trained on full-frequency inputs exhibited poor generalization (red curve), the low-to-full model (green curve) yielded high generalization, performing better than all other regimens across nearly the entire range of test frequencies. This is particularly noteworthy given that the full-frequency inputs subsume the entire low-pass-filtered content, and considering that the exclusively-full network could have learned to discard high frequencies were they not required to achieve high performance levels on the task. Generalization curves from the remaining two conditions (black curve for the exclusively-low model; blue curve for the full-to-low model) are consistently lower than that of the biomimetic one. Nevertheless, they indicate that including low-pass-filtered inputs into some phase of the training enhances performance over the exclusively full-frequency training regimen.

To probe the relationship between performance scores and the receptive field characterizations described earlier, we assessed classification performance on full-frequency inputs while gradually ablating the individual networks' filters with the highest peak frequencies. While the network trained on exclusively full-frequency inputs exhibited the largest performance decrement, the low-to-full model was least affected by the removal of higher frequency units, which is noteworthy given its spectral similarity to the networks trained using the full-to-low and exclusively-low regimens (see Figure 2.4b). Moving beyond the first network layer, we also examined the response stability of the deeper networks' units when presented with full-frequency versus low-frequency (500 Hz) test set inputs, operationalized as correlation score (r). As depicted in Figure 2.4c, across all four convolutional layers, activations in the network trained on full-frequency inputs were most varied, while the low-to-full network showed the least variation, further attesting to its invariance to high-frequency modulations.

2.5 DISCUSSION

Taken together, the computational results presented in this paper lend support to the proposal that commencing auditory development with degraded (i.e., exclusively low-frequency) stimuli may help set up temporally extended processing mechanisms that remain in place throughout subsequent experience with full-frequency inputs. The

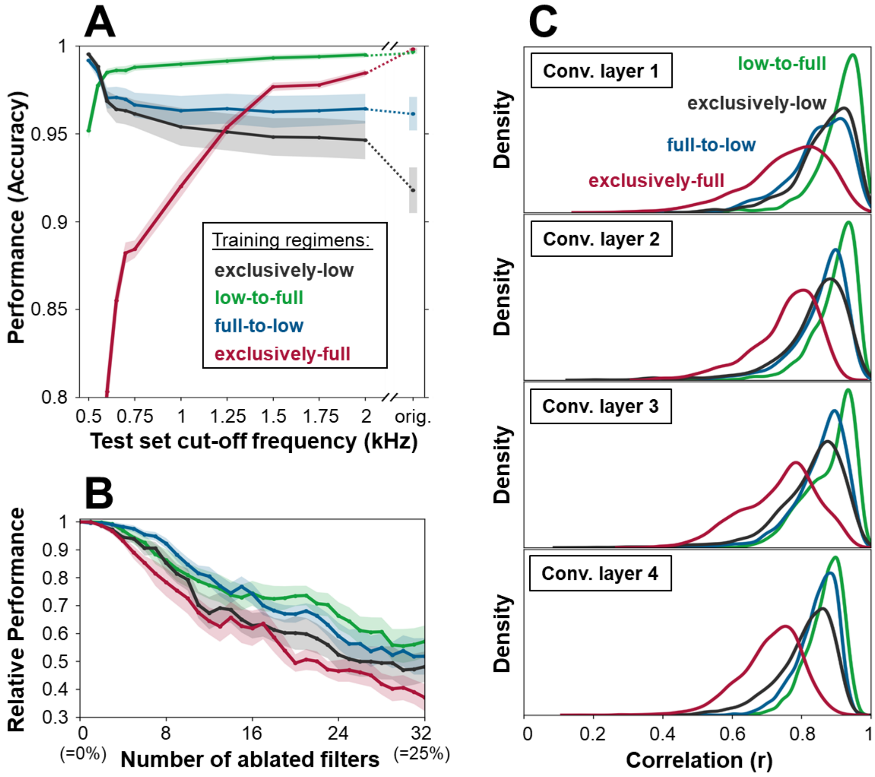


Figure 2.4: (a) Mean and standard error of 10-fold cross-validated emotion recognition performances on full-frequency and various low-frequency test sets. (b) Mean and standard error of 10-fold cross-validated classification performances (baseline-normalized, for visualization), when gradually removing the units with highest peak frequency. (c) Distribution of correlations of units' activities between low-pass-filtered and full-frequency inputs, across layers (10-fold pooled)

early instantiation of such mechanisms may, in turn, facilitate robust analyses of long-range modulations in the auditory stream, as indicated by generalized performance of the developmentally inspired model when tested on emotion recognition in full- and low-frequency conditions.

Beyond the superior generalization on the temporally extended task of emotion recognition, the observed robustness of classification performance to low-pass filtering not only mimics an important feature of auditory recognition in humans (Bornstein et al., 1994) but might also facilitate auditory analyses in more challenging environments. For instance, as sound travels through air, its energy is more substantially reduced in the higher, compared to the lower, frequencies (Little et al., 1992). One of the benefits that facility with using low frequencies could thus confer to the mature auditory system is robust classification even over large distances. Along similar lines, such proficiency may also be responsible for the ability of patients suffering from age-related hearing loss, which first affects the sensitivity of higher-frequency bands, to continue to be able to identify speech sounds for part of the progression of their hearing loss (Gates and Mills, 2005).

Our results suggest that precluding early experience with low-frequency inputs leads to a reduced emphasis on receptive fields tuned to low frequencies, with corresponding decrements in prosodic classification performance, while not necessarily affecting the processing of informational content in speech that is based on higher frequencies. Interestingly, this computationally derived result is corroborated by clinical data. Prematurely born infants, who did not get to experience exclusively low-frequency sounds for as long as normal-term infants but were almost immediately immersed into a full-frequency environment (roughly akin to the exclusively-full regimen in the computational simulations presented in this paper), have been reported to later exhibit impairments in the processing of low-frequency structure of sounds and show lower performance on prosody and emotion classification, even though their punctate acoustic detection thresholds are near normal (Amin et al., 2015; Gonzalez-Gomez and Nazzi, 2012; Ragó et al., 2014). This result has implications for the design of auditory environments for preterm babies in classical neonatal ICUs, which have been demonstrated to expose infants to frequencies above 500 Hz for the majority of the time (Lahav, 2015). Auditory interventions in neonatal ICUs are, if applied, usually oriented towards silence (Altuncu et al., 2009; Milette, 2010) or musical sounds (De Almeida et al., 2020; Loewy et al., 2013; Lordier et al., 2019) and have been reported to yield some benefits in terms of higher-level functioning. The findings presented in this paper, however, suggest that for preterm babies, the auditory environment in a neonatal ICU, instead of, or in addition to, being controlled for the general type of audio signal, should be made more similar to the intrauterine one in terms of the specific spectral composition of sounds, thereby mirroring, and adding support to, previous perspectives on the impact of high-frequency noise in neonatal ICUs (Lahav, 2015; Lahav and Skoe, 2014). Specifically, based on the data we presented in this paper, we argue that the environmental sounds allowed to reach the infant should be filtered with an approximation of the intrauterine acoustic filtering characteristics.

It remains to be seen how the converging computational and empirical evidence on the benefits of initial exposure to exclusively low-frequency sounds early in development, as derived as part of the computational results presented in this paper in the context of emotion recognition, generalizes to other ecologically relevant aspects and tasks of hearing such as the comprehension of other prosodic content. This could, in the future, be probed in preterm versus full-term babies and possibly be further examined as a function of the auditory intervention applied to the neonatal ICU environment.

Finally, it is worth noting that the data presented here are consistent with results previously reported in the visual domain (Vogelsang et al., 2018). There it was found that initially low spatial acuity, that progressively improved over the first years of life, helped instantiate receptive fields capable of extended spatial integration, and more robust classification performance. The analogous results we have found in the auditory domain suggest that the adaptive benefits of initial sensory limitations may apply across different modalities. Thus, in addition to highlighting the potential benefits of sound degradation in the intrauterine environment, the results presented in this paper may also help explain the adaptive significance of some key human developmental progressions, prenatal or postnatal, and across sensory domains. The initially degraded inputs may provide a scaffold rather than act as hurdles. Taken together, this may not only help elucidate some of the mechanisms underlying our later auditory proficiencies, but also help us better understand why the choreography of developmental stages is structured in the way that it is. From the perspective of artificial intelligence, the salutary effects of these developmental trajectories upon later classification performance of biological systems suggest that computational systems for analogous tasks may also benefit from incorporating biomimetic training regimens.

acknowledgements

This work was supported by NIH Grant R01EY020517.

conflict of interest

The authors declare no conflicts of interest.

ETHICS STATEMENT

No human data were collected as part of this study.

DATA AVAILABILITY STATEMENT

The computational model code is openly available at https://github.com/marin-oz/PrenatalAudition.

2.6 SUPPORTING INFORMATION

2.6.1 *Extended Methods*

2.6.1.1 *Network architecture, parameters, and training procedure*

We utilized the "M5" model by Dai et al. (2017), equipped with 4 convolutional layers (each involving convolution, batch normalization, application of the ReLU activation function, and max pooling), followed by global average pooling, and connected to the output nodes of the network, representing the 7 different emotion classes, through a single dense layer with softmax activation function. The number of units in the different layers was taken from Dai et al. (2017), with the exception of the kernel size in the first convolutional layer, which, due to a different sampling frequency in the dataset (24414 Hz), was adjusted to 244, to span a 10ms window of the auditory input – a time window frequently used in MFCC (Mel Frequency Cepstral Coefficients)-based computational audition models – as suggested by the authors of the computational model used (Dai et al., 2017). The network was implemented in Keras and trained on a single GPU, using stochastic gradient descent with a batch size of 32 and a standard learning rate of 0.01. All four regimens were trained for a total of 100 epochs ('low-to-full' for 50 epochs on low-frequency and 50 epochs on full-frequency inputs, 'full-to-low' for 50 epochs on full-frequency and 50 epochs on low-frequency inputs, 'exclusively-full' for 100 epochs on full-frequency inputs, and 'exclusively-low' for 100 epochs on low-frequency inputs). The number of epochs was chosen to ensure that the regimens had converged and showed little variation between runs, to ensure comparability and stability of results.

2.6.1.2 *Analysis of representations and activation*

For the receptive field analysis reported in Figures 2.2 and 2.3, we examined the filters in the first convolutional layer of the networks by subjecting them to the Fourier transform. Figures 2.2A-B depict the power values extracted for frequencies between 0.1 and 6 kHz (the maximum we chose), in steps of 0.1 kHz, with the sum up to 6 kHz normalized to 1. Note that these frequency values, and the ones depicted on the axes in Figure 2.2, are rounded: With a frequency increment of 24414/244, the rounded frequency values (0.1, 0.5, 1, 2, 3, 4, and 6 kHz) correspond to 100.0574, 500.2869, 1000.5738, 2001.1475, 3001.7213, 4002.2951, and 6003.4426 Hz. For the correlational analysis reported in Figure 2.4C, for each trained network and each of the 10 folds, correlations were computed between the activations of units, concatenated for all 280 test set items, following full-frequency vs. low-frequency inputs. Figure 2.4C depicts the distribution of these individual units' correlations.

REFERENCES

Altuncu, E, I Akman, S Kulekci, F Akdas, Hülya Bilgen, and E Ozek (2009). "Noise levels in neonatal intensive care unit and use of sound absorbing panel in the isolette." In: *International journal of pediatric otorhinolaryngology* 73.7, pp. 951–953.

Amin, Sanjiv B, Mark Orlando, Christy Monczynski, and Kim Tillery (2015). "Central auditory processing disorder profile in premature and term infants." In: *American journal of perinatology* 32.04, pp. 399–404.

Bornstein, Steven P, RH Wilson, and Nancy K Cambron (1994). "Low- and high-pass filtered Northwestern University Auditory Test No. 6 for monaural and binaural evaluation." In: *Journal of the American Academy of Audiology* 5.4, pp. 259–264.

Dai, Wei, Chia Dai, Shuhui Qu, Juncheng Li, and Samarjit Das (2017). "Very deep convolutional neural networks for raw waveforms." In: *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, pp. 421–425.

De Almeida, Joana Sa, Lara Lordier, Benjamin Zollinger, Nicolas Kunz, Matteo Bastiani, Laura Gui, Alexandra Adam-Darque, Cristina Borradori-Tolsa, François Lazeyras, and Petra S Hüppi (2020). "Music enhances structural maturation of emotional processing neural pathways in very preterm infants." In: *Neuroimage* 207, p. 116391.

Dominguez, Melissa and Robert A Jacobs (2003). "Developmental constraints aid the acquisition of binocular disparity sensitivities." In: *Neural Computation* 15.1, pp. 161–182.

Dupuis, Kate and M Kathleen Pichora-Fuller (2010). *Toronto emotional speech set (TESS)*.

Elman, Jeffrey L (1993). "Learning and development in neural networks: The importance of starting small." In: *Cognition* 48.1, pp. 71–99.

Gates, George A and John H Mills (2005). "Presbycusis." In: *The lancet* 366.9491, pp. 1111–1120.

Gerhardt, Kenneth J and Robert M Abrams (1996). "Fetal hearing: characterization of the stimulus and response." In: *Seminars in perinatology*. Vol. 20. 1. Elsevier, pp. 11–20.

Gonzalez-Gomez, Nayeli and Thierry Nazzi (2012). "Phonotactic acquisition in healthy preterm infants." In: *Developmental science* 15.6, pp. 885–894.

Griffiths, Scott K, WS Brown Jr, Kenneth J Gerhardt, Robert M Abrams, and Richard J Morris (1994). "The perception of speech sounds recorded within the uterus of a pregnant sheep." In: *The Journal of the Acoustical Society of America* 96.4, pp. 2055–2063.

Hepper, Peter G and B Sara Shahidullah (1994). "Development of fetal hearing." In: *Archives of Disease in Childhood - Fetal and Neonatal Edition* 71.2, F81–F87. ISSN: 1359-2998.

Lahav, Amir (2015). "Questionable sound exposure outside of the womb: frequency analysis of environmental noise in the neonatal intensive care unit." In: *Acta paediatrica* 104.1, e14–e19.

Lahav, Amir and Erika Skoe (2014). "An acoustic gap between the NICU and womb: a potential risk for compromised neuroplasticity of the auditory system in preterm infants." In: *Frontiers in neuroscience* 8, p. 381.

Little, Alex D, Donald H Mershon, and Patrick H Cox (1992). "Spectral content as a cue to perceived auditory distance." In: *Perception* 21.3, pp. 405–416.

Loewy, Joanne, Kristen Stewart, Ann-Marie Dassler, Aimee Telsey, and Peter Homel (2013). "The effects of music therapy on vital signs, feeding, and sleep in premature infants." In: *Pediatrics* 131.5, pp. 902–918.

Lordier, Lara, Djalel-Eddine Meskaldji, Frédéric Grouiller, Marie P Pittet, Andreas Vollenweider, Lana Vasung, Cristina Borradori-Tolsa, François Lazeyras, Didier Grandjean, Dimitri Van De Ville, et al. (2019). "Music in premature infants enhances high-level cognitive brain networks." In: *Proceedings of the National Academy of Sciences* 116.24, pp. 12103–12108.

Milette, Isabelle (2010). "Decreasing noise level in our NICU: the impact of a noise awareness educational program." In: *Advances in Neonatal Care* 10.6, pp. 343–351.

Murkoff, Heidi (2016). *What to expect when you're expecting*. Workman Publishing.

Newport, Elissa L (1988). "Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language." In: *Language sciences* 10.1, pp. 147–172.

Norman-Haignere, Sam V and Josh H McDermott (2018). "Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex." In: *PLoS biology* 16.12, e2005127.

Ragó, Anett, Ferenc Honbolygó, Zsófia Róna, Anna Beke, and Valéria Csépe (2014). "Effect of maturation on suprasegmental speech processing in full-and preterm infants: A mismatch negativity study." In: *Research in developmental disabilities* 35.1, pp. 192–202.

Ross, Mark, Robert J Duffy, Harry S Cooker, and Russell L Sargeant (1973). "Contribution of the lower audible frequencies to the recognition of emotions." In: *American Annals of the Deaf*, pp. 37–42.

Snel, John and Charlie Cullen (2013). "Judging emotion from low-pass filtered naturalistic emotional speech." In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, pp. 336–342.

Turkewitz, Gerald and Patricia A Kenny (1982). "Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement." In: *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 15.4, pp. 357–368.

Vogelsang, Lukas, Sharon Gilad-Gutnick, Evan Ehrenberg, Albert Yonas, Sidney Diamond, Richard Held, and Pawan Sinha (2018). "Potential downside of high initial visual acuity." In: *Proceedings of the National Academy of Sciences* 115.44, pp. 11333–11338.

Webb, Alexandra R, Howard T Heller, Carol B Benson, and Amir Lahav (2015). "Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation." In: *Proceedings of the National Academy of Sciences* 112.10, pp. 3152–3157.

3 IMPACT OF EARLY VISUAL EXPERIENCE ON LATER USAGE OF COLOR CUES

Content from

Vogelsang, M.*, Vogelsang, L.*, Gupta, P.*, Gandhi, T., Shah, P., Swami, P., Gilad-Gutnick, S., Ben-Ami, S., Diamond, S., Ganesh, S., & Sinha, P. (Submitted). "Impact of early visual experience on later usage of color cues". **Under peer-review** at Science (initial submission: September 2023). (* = equal contribution)

3.1 ABSTRACT

Human visual recognition is remarkably robust to chromatic changes. Here we offer a potential account of the roots of this resilience based on observations with ten congenitally blind children who gained sight late in life. Several months or years following their sight-restoring surgeries, the removal of color cues significantly reduced their recognition performance while age-matched controls showed no such decrement. This finding may be explained by the greater-than-neonatal maturity of the late-sighted children's color system at sight onset, inducing overly strong reliance on chromatic cues. Results of simulations with deep neural networks corroborate this hypothesis. These findings offer an account of why color usage is impacted by early deprivation, highlight the adaptive significance of typical developmental trajectories, and provide guidelines for enhancing machine vision systems.

3.2 INTRODUCTION

Catarrhine primates, including humans, are endowed with excellent color vision, perhaps second only to avian species, in terms of the spectral breadth and resolution of chromatic information they experience (Jacobs, 2008). Likely driven by evolutionary pressures (Carvalho et al., 2017; Dominy and Lucas, 2001; Regan et al., 2001), they are the only group among placental mammals to possess trichromatic vision (Kawamura, 2016). The eminence of color processing in primates is also reflected in the underlying neurophysiology. The primary visual cortex exhibits strong color sensitivity (Hurlbert, 2003), and nearly two-thirds of the neurons therein are color-selective (Shapley and Hawken, 2011). The preponderance of color-tuned units is maintained and even enhanced as one progresses along the ventral visual stream (Kozlovskiy and Rogachev, 2021), believed to play a significant role in recognition (Righi and Vettel, 2011).

In light of the prominence of color in perceptual experience and neurophysiology, it is surprising that our ability to recognize objects in images that are devoid of color registers no significant decrement relative to what is feasible with full-color images (Biederman and Ju, 1988; Davidoff and Ostergaard, 1988; Elder and Velisavljević, 2009; Mapelli and Behrmann, 1997; Marx et al., 2014; Meng and Potter, 2008). This is evidenced by the ease with which we are able to recognize people and objects in old black-and-white photographs and movies. What accounts for this robustness to color desaturation and broad generalization across color shifts?

Our facility with monochrome images could potentially be attributed to our having had prior access to these kinds of stimuli as pictures in newspapers and books; a form of learning that establishes equivalence between natural color imagery and its depiction in grayscale pictures. This notion of grayscale image recognition as a manifestation of explicit training with such stimuli has been tested using a variety of approaches. Notably, Hochberg and Brooks (1962) experimented with their own son to test this hypothesis, rearing him without access to any pictorial material whatsoever. Given the challenges inherent in conducting the study, they stopped the deprivation regimen when the boy reached 19 months of age. Impressively, the child was able to recognize grayscale pictures at near-ceiling level the very first time he saw them. This finding has since been replicated with many more infants (DeLoache et al., 1979; Shinskey and Jachens, 2014). Additional evidence stems from anthropological studies of remote tribe members that show unimpaired identification of grayscale pictures, despite the absence of prior exposure (Deregowski et al., 1972). While leaving unanswered precisely how the high recognition performance with grayscale images is achieved, these studies provide compelling evidence that this ability is not a cultural artifact and does not depend on explicit training with such imagery.

3.3 EMPIRICAL STUDIES WITH NEWLY SIGHTED INDIVIDUALS

The work reviewed above makes a prediction regarding the results expected from individuals who gain sight late in life. Given the non-necessity of exposure to grayscale images for resilience to chromatic changes, for the late-sighted, too, performance with color and grayscale images should be equivalent. We decided to test this hypothesis with children we have been working with as part of a humanitarian and scientific initiative named Project Prakash (Sinha, 2013). This effort identifies children with treatable congenital blindness and provides them sight surgeries, subsequently affording the opportunity to study their visual development and to develop new methods to enhance it. Interestingly, and as detailed below, we found that the hypothesis of color and grayscale equivalence was not supported by the Prakash data. This has pointed us to a hypothesis that not only provides a possible explanation for the empirical data from the Prakash children but also a potential account for why typically sighted individuals exhibit resilience to color changes.

In our first experiment, we tested ten early-blind Prakash individuals aged 7-26 years. All had dense bilateral cataracts detected at, or within six months of, birth. Assessment of visual history was based on parental reports, ophthalmic examination of ocular structures, and the presence of nystagmus, which is known to be induced by profound visual impairment very early in life (Tusa et al., 1991). The patients were provided surgeries, which involved cataract extraction and intraocular lens implantation, and were tested several months or years thereafter. The detailed patient profiles are included in Supplemental Table 3.1. The control group comprised 10 normally sighted age- and socio-economic status-matched children. They were tested while wearing blur goggles simulating Snellen visual acuities of 20/200 and 20/500 (non-overlapping groups of 5 each), to bracket the range of Prakash patients' acuities and approximately match their average acuity. Blurring was achieved by attaching Bangerter occlusion foils (Odell et al., 2008) to clear safety goggles. By assessing the performance of the normally-sighted while wearing blurring goggles, we titrated the effects of reduced acuity, comparable to the Prakash children's post-operative outcomes, from non-optical factors on image classification performance. There was no history of neurological or psychiatric illness in any of the participants. All experiments were approved by the IRBs of MIT and Dr. Shroff's Charity Eye Hospital – our medical partner in New Delhi.

In the experiment, participants were asked to name commonplace living/non-living objects that were presented in color (see Figure 3.1A, left) and in grayscale (see Figure 3.1A, right). A total of 100 images were included in the study. First, monochrome images were shown, one at a time, and participants were asked to name the presented object. Then, in a second session, full-color images were presented. No time limits were placed on the participants' examination of each stimulus. All images were displayed on a 21-inch computer monitor and viewed from an average distance of 40 cm, subtending 60 degrees of visual angle horizontally. The results reveal a highly significant difference between the Prakash and control group in classification performance on grayscale images ($t(18) = 3.979$, $p < .001$ in two-tailed independent t-test), when normalizing performance on color images to 100%, in order to account for individual differences in absolute recognition capabilities (Figures 3.1B&C).

These results present us with two questions: How do normally sighted achieve full generalization, and what prevents Prakash children from doing so? We present a hypothesis based on early developmental trajectories and describe computational results that corroborate

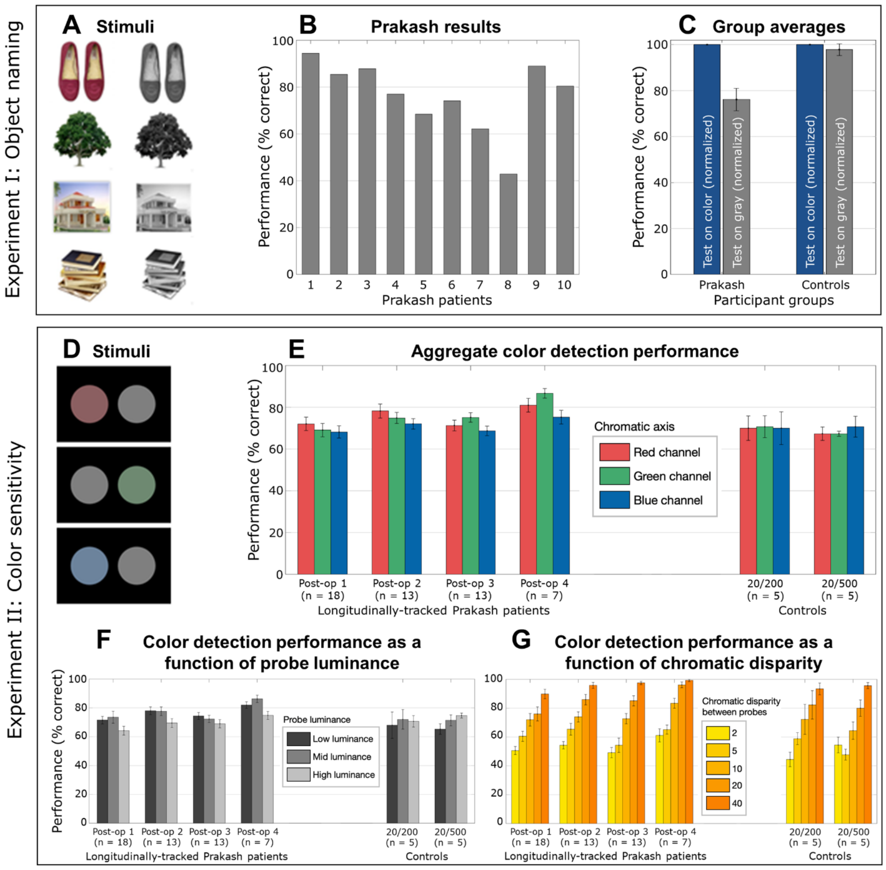


Figure 3.1: **A.** Sample stimuli for the object naming experiment in the color (left) and grayscale condition (right). **B.** Naming results of Prakash children, depicting individual participant data on the recognition of grayscale objects, when normalizing recognition performance on full-color images to 100%. Performance was consistently lower when color information was removed from the images. **C.** Performance means of Prakash children and blur-matched controls, on color and grayscale images, revealing a highly significant group difference ($t(18) = 3.979$, $p < .001$) in normalized grayscale performance. Error bars depict the standard error. **D.** Exemplar stimuli for the color sensitivity experiment. In each trial, a pair of discs was presented, one gray and the other colored. Participants were asked to indicate which disc was colored. **E.** Results of the color sensitivity test, displayed as a function of color and depicted for Prakash patients at four different time points (post-op 1: ca. 2 days after surgery; post-op 2: ca. 7 days after surgery; post-op 3: ca. 30 days after surgery, and post-op 4: ca. 6 months after surgery) as well as controls wearing 20/200 and 20/500 blur goggles. **F.** Results of the color sensitivity test, when displayed as a function of luminance intensity. **G.** Results of the color sensitivity test, when displayed as a function of delta (i.e., the difference in hue between the two discs). Error bars in panels E-G depict the standard error.

it. This hypothesis builds on the observation that typically developing infants start with immature retinas; more specifically, the cone photoreceptors in the neonatal retina are significantly limited in transduction capabilities relative to their mature counterparts (Yuodelis and Hendrickson, 1986). This immaturity compromises neonatal color vision, reducing the chromatic content of inputs that the infant experiences (Skelton et al., 2022; Teller, 1998). By contrast, the Prakash children commence their post-operative vision with mature cone cells that subserve near-normal color vision. This is borne out by the results of tests of color sensitivity we conducted as part of a second experiment with Prakash children.

In our second experiment, we tested 18 Prakash patients (aged 8-20 years; 7 females) immediately and longitudinally following surgeries for bilateral congenital cataracts (see Supplemental Table 3.2 for detailed patient profiles) as well as 10 normally-sighted controls (mean age 15.7; 7 females). Participants were shown pairs of discs presented on a black screen (Figure 3.1D). While one of the discs was gray, the other disc was colored (the amount of hue was controlled by a parameter corresponding to the difference in the R, G or B values between the two discs). For each of 90 trials, participants were asked to point to the disc they perceived to be colored. As in the previous task, control participants performed the test while wearing blurring goggles, with induced blur corresponding to 20/200 and 20/500 Snellen acuity. As is evident in Figure 3.1E, even in the first post-operative session, conducted just two days post-surgically, the Prakash children achieve accuracies statistically indistinguishable from normally sighted controls for each of red, green, and blue colors (no significant group differences ($p = 0.920$, $\eta^2 = 0.002$) or <group x color channel> interactions ($p = 0.943$, $\eta^2 = 0.010$) in a 2-way ANOVA; Supplemental Table 3.3). Similarly, examining the data as a function of the luminance of the discs (Figure 3.1F) or the difference in hue between the discs (Figure 3.1G), Prakash children and controls perform remarkably similarly. These results indicate that the Prakash children have a mature, well-functioning color system right at the outset of their post-operative vision.

Although good color vision from the start might intuitively appear to be a desirable feature for the visual system, we hypothesize that immediate immersion of the Prakash children into color-rich imagery, rather than following a gradual progression from color-deficient to color-rich, may, in fact, be detrimental. Such immersion may induce an unnaturally strong reliance on color cues. In contrast, for the normally-sighted, who underwent typical developmental trajectories, early experience with color-degraded inputs may prove to be beneficial by instantiating representations that emphasize luminance rather than chromatic cues, thereby having implicit resilience to color removal.

Our hypothesis, thus, is that robustness to color removal is the consequence of the normal developmental progression from impoverished to rich color vision. Eliminating the initial phase of this progression, as it happens with Prakash children, leads to an overly strong reliance of image representations on color cues, with attendant drops in performance when color information is removed. This hypothesis has the potential to answer both of the questions we introduced earlier: How do normally sighted observers achieve generalization to grayscale imagery, and why do the late sighted have trouble doing so?

3.4 COMPUTATIONAL STUDIES WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

To test the above hypothesis systematically, we need to be able to manipulate the temporal progression of sensory experience and examine the consequences of such manipulation. While this is ethically and practically infeasible in human participants, studies with computational model systems, capable of directly learning from experience, offer a way forward. Here, we used deep convolutional neural networks (DCNNs), which are among the most successful in predicting human behavior and neural responses across the sensory hierarchy (Cadena et al., 2019; Lindsay, 2021; Schrimpf et al., 2020; Storrs et al., 2021) and allow for a direct assessment of the impact of different training protocols (or 'regimens') on the system's later performance and learned representations. Some of these regimens are thereby chosen to be biomimetic, in the sense that they recapitulate aspects of normal developmental trajectories, while others serve as non-developmental controls.

For the results reported in the main manuscript, we utilized the well-established and compact AlexNet architecture (Krizhevsky et al., 2012), with relatively simple parameter settings featuring, e.g., constant learning rates, in order to render subsequent analyses most interpretable. In addition, we examined the generalizability of our results as a function of architecture (specifically, using the Resnet-50 (He et al., 2016) and Inception v3 (Szegedy et al., 2016) networks), parameter settings (specifically, batch size and learning rate), as well as choices of learning rate schedules and normalization, as reported in Supplemental Figures 3.4-3.8.

To examine the consequences of exposure to inputs undergoing several temporal progressions of chromaticity, we trained different instances of our network on the ImageNet (Deng et al., 2009) and FaceScrub (Ng and Winkler, 2014) databases while controlling the availability of color information throughout training. Specifically, we trained on two regimens that, while not mimicking all of the details, correspond to the general transition from color-degraded to color-rich experience in typical development as well as to experience with full-color imagery immediately upon late sight onset:

- 'Gray-to-color' (G2C), in which, as a coarse proxy for the developmental trajectory of typically developing infants, initial training on grayscale images (on half of the total number of epochs) is followed by later training on color images (on the other half)
- 'Color-to-color' (C2C), in which, roughly akin to the experience of Prakash children, the entire training set consists of full-color images

For completeness, we included two additional regimens in our tests: 'Gray-to-gray' (G2G), in which the entire training consists of grayscale images, and 'Color-to-gray' (C2G) in which training follows an inverse-developmental progression but has the same aggregate content as that of the G2C regimen. This would reveal the significance, if any, of the temporal sequencing of experience, rather than simply the aggregate composition of the training set.

An analysis of the resulting ImageNet classification performances revealed that while C2C training resulted in high performance on color but poor performance on grayscale images, the G2C training regimen yielded stable performance levels for both (Figure 3.2A). Notably, the C2G network, which had been trained on the same aggregate content as the G2C network, only in reversed temporal order, exhibited markedly inferior generalization. These results, revealing a salient ordering effect, suggest that commencing training with initially color-degraded inputs, in keeping with normal development, may confer benefits to later generalization, compared to training that commences with colored imagery. These benefits, furthermore, also apply to other chromatic variations, such as hue rotations, by which the G2C model, unlike models initially trained on color inputs (C2C or C2G), is almost entirely unaffected (Figure 3.2B).

These findings highlight the benefits of the G2C model, a rough proxy for typical development, over the C2C model, a rough proxy for the experience of Prakash children. We further rule out the possibility that the benefits of the G2C model can be reduced to data augmentation, considering that the C2G model, which has been trained on the same overall content as the G2C model but in reverse order, performs especially poorly. What is not evident from the data presented thus far, however, is what benefits the G2C regimen may have over a G2G regimen, considering the similarity in performance of the two (Figures 3.2A and 3.2B). This relates to the broader question of the adaptive significance of including chromatic sensitivity in visual systems. First, the benefits of color vision have primarily been found outside of the domain of recognition, including, for instance, deciding a person's gender, or assessing their health based on facial hues, rather than recognizing that same person (Hiramatsu et al., 2017; Jones, 2018; Nestor

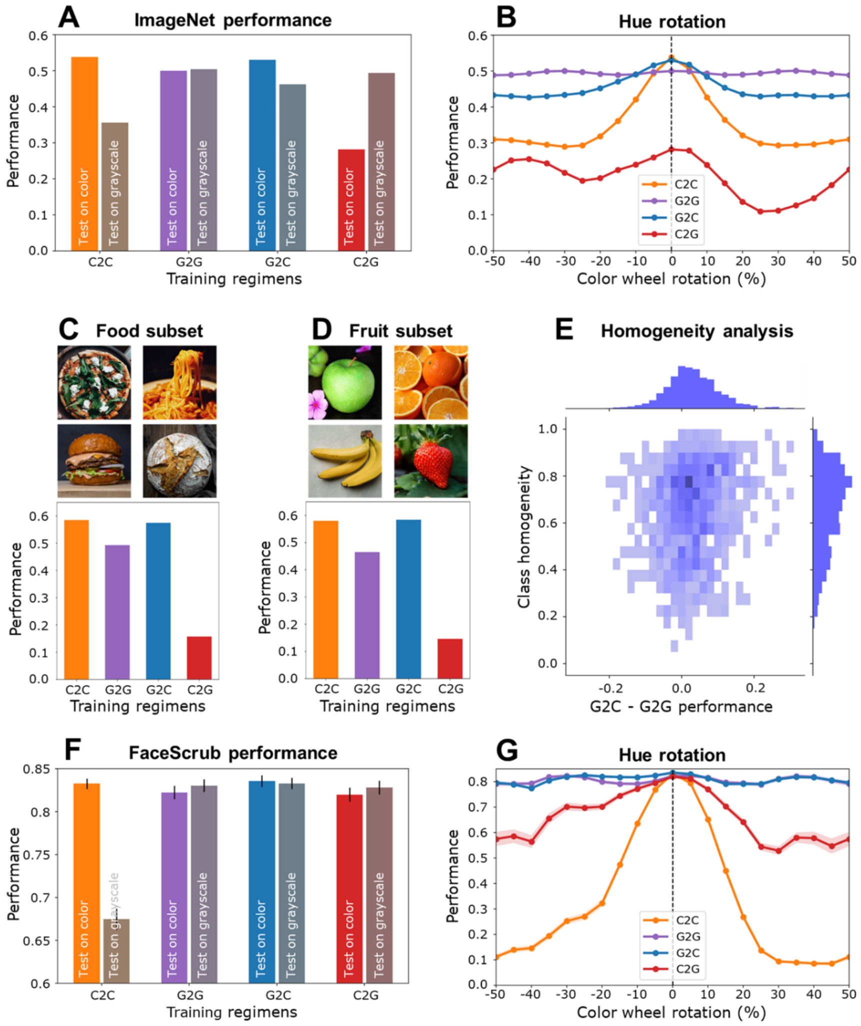


Figure 3.2: **A.** Color and grayscale classification performance of networks trained on the ImageNet database using our four different regimens. **B.** Classification performance of different networks tested on the ImageNet database when rotating the color wheel by gradually shifting the hue content of the original image. **C & D.** Color classification performances of networks trained on the ImageNet database, evaluated on the subset of classes depicting food (C) and fruit (D) items. **E.** Relationship between class-specific performance gains of the G2C model over the G2G model, and the homogeneity of mean hues across the different instances of a given class (see Supplemental Methods for details). **F.** 10-fold cross-validated classification performances of the four networks trained on the FaceScrub database. Error bars depict the standard error. **G.** 10-fold cross-validated classification performances of networks trained and tested on the FaceScrub database when rotating the color wheel. The shaded area depicts the standard error.

and Tarr, 2008; Stephen et al., 2009; Thorstenson et al., 2020). Thus, having a G2C model capable of performing at least as well as the G2G model, while also being equipped with color sensitivities required for analyses in other domains, would be beneficial. Second, even within the domain of visual recognition, specific sets of classes, such as fruits or food, have been found to particularly benefit from color sensitivity (Melin et al., 2019; Nevo et al., 2018; Spence, 2018).

To test whether results of our computational simulations would support these qualitative suggestions, we examined classification performance on a subset of ImageNet classes depicting food items and, as a further subset thereof, fruit classes. Our results confirm that for recognition of colored imagery in these domains, the G2C model is indeed superior to the G2G model, reaching performance levels as high as that of the C2C model (Figure 3.2C&D). Further, examining the relationship between the performance gains of the G2C model relative to the G2G model on the one hand, and the homogeneity of mean hues across the different instances of a given class on the other, one can observe a subset of classes that particularly benefits from the G2C model and exhibits relatively high hue homogeneity (Figure 3.2E). However, due to factors such as variations in background, this relationship is an imperfect one.

Similar to the basic results observed in Figure 3.2A&B, also when trained and tested on images of faces, the G2G and G2C models were the only ones that simultaneously achieved strong generalization across color, grayscale, and hue rotation conditions, with the G2C network even reaching slightly higher performance levels for the colored test sets (Figure 3.2F&G). It is important to note that these general patterns of results hold not only across databases but also across architectures, learning rates, and other parameter choices (Supplemental Figures 3.4-3.8).

Next, to study potential mechanistic underpinnings of the observed differences in generalization towards color removal and hue shifts, we analyzed the learned representations of our trained networks. An inspection of the individual receptive fields in the first convolutional layers of the networks following uniform training revealed that while the C2C network is equipped with a mix of colored and uncolored filters, the G2G network, as expected from its inputs, possesses exclusively achromatic filters (Figure 3.3A). Interestingly, we found that when color was followed by grayscale inputs in the second phase of training, the receptive fields subsequently exhibited a strong reduction of color content (see C2G in Figure 3.3B) but that when grayscale was followed by color inputs, only some of the receptive fields showed a corresponding gain in color content (see G2C in Figure 3.3B). Quantifying this observation, a 2-sample t-test revealed a highly significant difference in the absolute amount of color change between the two networks ($t(190) = 10.80$; $p < .001$). This effect of ordering suggests that achromatic receptive fields may constitute a more robust starting point for the front end of a visual recognition system during training.

To examine deeper-layer representations, we synthesized images eliciting maximal unit activation, following Olah et al. (2017), for each of the layers and each of the four trained networks. Sets of exemplar synthesized images for a middle layer (the fourth convolutional layer), as well as the last fully-connected layer, are depicted in Figures 3.3C&D (for additional visualizations, see Supplemental Figures 3.9 and 3.10). Similar to the receptive fields (Figure 3.3A), the synthesized images for the G2C model appear to exhibit structural similarity with those of the model trained exclusively on grayscale imagery (G2G), with the former showing a subtle colorization of the latter. Interestingly, visual inspection of the synthesized images for the last layer of the G2C model suggests that this rather modest addition of chromatic information may facilitate classification decisions, particularly for fine-grained discrimination, while, at the same time, retaining much of the structural information present in the synthesized images of the G2G model. For the C2C model, by contrast, synthesized images appear to contain weaker structural features, inducing a greater reliance on specific color cues for classification. Finally, the synthesized images for the C2G model exhibit significantly more ambiguous structural and chromatic cues, possibly due to the generally low classification performance (Figure 3.2A).

The recognizability of these synthesized images provides a measure of the class-specific information they capture. Accordingly, we conducted an online experiment to examine it. A total of 39 participants were presented with synthesized images belonging to one of three domains – food, animals, or other objects, with each of these superordinate categories comprising 32 basic-level classes. The G2G, G2C, and C2C models' synthesized images were shown in random order, and participants were asked to classify each image into (i) a super-category (by clicking on one of the labels "food", "animals", or "objects") and, subsequently, (ii) into one of the 32 basic-level classes belonging to the chosen super-category, whose labels were presented on the screen (for detailed experimental procedures, see Supplemental Methods). We did not include the C2G model's synthesized images as they appeared entirely unrecognizable.

The results of this study demonstrate that participants performed significantly better in super-category classification, averaged across the three classes, for the synthesized images of the G2C model relative to those of the G2G model ($t(38) = 2.49$; $p = 0.017$), for which itself performance was significantly higher than for the C2C model ($t(38) = 5.583$; $p < 0.001$) (Figure 3.3E). When examining basic-level classification performance, the superiority of recognizing the G2C over the G2G model's images was further emphasized, with the former exceeding the latter by 10.7% percentage points (Figure 3.3F) and

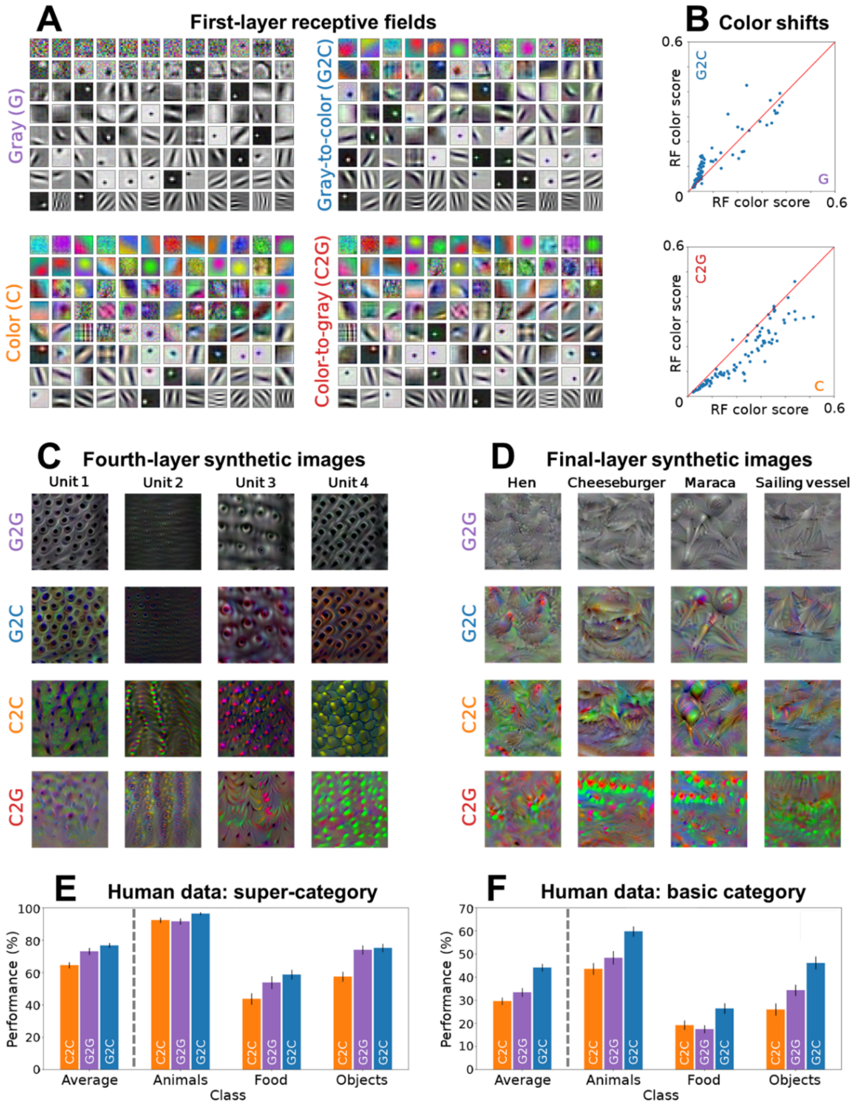


Figure 3.3: **A.** Depiction of the 96 individual receptive fields of our four networks trained on the ImageNet database, sorted by their colorfulness (see Supplemental Methods for details). **B.** Comparison of the colorfulness of individual filters when transitioning from the first to the second stage of training, depicted separately for the transition of gray to gray-to-color as well as color to color-to-gray. **C&D.** Depiction of exemplar synthesized images (eliciting maximal unit activation) from the fourth convolutional (C) and last fully-connected (D) layer, for our four different training regimens. **E&F.** Performance (in % correct) of n = 39 online participants in classifying synthesized images from the G2G, G2C, and C2C models into one of the three super-categories "Animals", "Food", or "Objects" (E) and classifying them into the correct super-category as well as the correct basic-level category (F).

highly significantly so ($t(38) = 9.24$; $p < .001$). Interestingly, in the basic-level classification task, performance on the C2C model's images is relatively close to that on the G2G model's images, leading to classification exceeding that for the G2G model in specific categories such as food classes (see Supplemental Figure 3.11 for basic-level performance conditioned on correct superordinate-category classification).

In sum, consistent with our hypothesis, the G2C model appears to benefit from having acquired stably strong luminance-based structural representations in the initial part of training with grayscale imagery, which are then supplemented with additional, subtle chromatic cues in the second part of training on colored inputs. To the contrary, training only on grayscale lacks the latter, training only on color lacks the former, and training on the inverse-biomimetic regimen leads to unstable representations and generalization behavior.

3.5 CONCLUSION

Taken together, the hypothesis and the computational results presented in this paper help account for two experimental findings – the robust generalization across color and color-shifted/desaturated images by normally-sighted observers, and the marked reduction in such ability by late-sighted individuals. We propose that initial limitations in color perception inherent in normal developmental progression may, in fact, be adaptive and advantageous, rather than represent a disadvantage. This provides a possible account for why normal development proceeds in the way that it does, and how deviations from it, even those that seemingly enhance input quality, can adversely impact performance. Interestingly, this hypothesis may also be relevant for understanding aspects of phylogenetic development (Jacobs et al., 2019). Besides furthering our understanding of normal development, this work also has important implications for designing clinical interventions. Specifically, in the context of late sight onset, the findings suggest that an initial period of deliberate color reduction immediately after surgery may be useful for facilitating later classification robustness. Additionally, these findings point to how incorporating insights from biological development can help improve generalization by machine vision systems.

The developmental perspective presented in this paper may also have relevance for understanding a key organizational principle of the mammalian visual system – the differential sensitivities of the magno- and parvocellular pathways. Specifically, the temporal confluence of poor acuity and low chromatic sensitivity early in development, and high acuity and rich color information later in the timeline, may help account for the analogous confluence in the response properties of the two neuronal streams (Livingstone and Hubel, 1988; Shapley, 1990). A separate report will describe this hypothesis and its tests in greater detail. Finally, it is relevant to note that we recently obtained results attesting to the potential benefits of the initially low visual acuity infants experience after birth (Vogelsang et al., 2018) as well as of low-pass filtered audio inputs as part of prenatal hearing (Vogelsang et al., 2023). Thus, the results we are presenting in this paper are, we believe, a special case of a broader set of phenomena that we refer to as 'Butterfly Effects', inspired by Lorenz's usage of this term for understanding the behavior of complex dynamical systems. In the context of biological development, this refers to the possibility that early perceptual limitations may be the subtle precursors that manifest in due time as significant salutary effects on perceptual skills.

acknowledgements

We wish to thank Dr. Bevil Conway and Dr. Anya Hurlbert for fruitful discussions. We wish to thank all of the participants who contributed to the study and their parents.

funding

This work was supported by grant R01EY020517 to PS.

author contributions

Conceptualization: Marin Vogelsang, Lukas Vogelsang, Sidney Diamond, Pawan Sinha. Data collection in India: Priti Gupta, Tapan Gandhi, Pragya Shah, Piyush Swami. Patient treatment and characterization: Suma Ganesh. Data analysis: Marin Vogelsang, Lukas Vogelsang, Priti Gupta, Sharon Gilad-Gutnick, Shlomit Ben-Ami, Pawan Sinha. Computational simulations: Marin Vogelsang, Lukas Vogelsang. Paper writing: Marin Vogelsang, Lukas Vogelsang, Priti Gupta, Sidney Diamond, Pawan Sinha. Review and final approval of the paper: All authors.

data and code availability

All data and code will be shared in a public repository.

competing interests

None of the authors declare competing interests.

3.6 SUPPLEMENTARY MATERIAL

3.6.1 *Supplementary Figures*

3.6.2 *Supplementary Tables*

3.6.3 *Supplementary Information and Methods*

3.6.3.1 *Prakash studies*

Patient information: The patient groups comprised early-onset blind individuals identified via rural outreach as part of Project Prakash. Except for one patient, all had dense bilateral cataracts since before one year of age (see Supplemental Tables 3.1 & 3.2). Assessment of congenitality of deprivation was based on parental reports, ophthalmic examination, and the presence of nystagmus, which is known to be induced by profound visual impairment very early in life (Tusa et al., 1991). The children were provided surgeries, which involved cataract extraction and intra-ocular lens implantation.

Pre-operative assessments: We tested for light perception in all four quadrants. The anterior segment was evaluated on a slit lamp, and the type of cataract and any associated ocular pathologies were noted. Given the patients' dense bilateral cataracts, which precluded fundus viewing via ophthalmoscopes, B scan ultrasound was carried out in all cases pre-operatively to check for any posterior segment pathology.

Intervention: Keratometry and biometry of all children was performed under general anesthesia just before the surgery. All surgeries were performed by a single surgeon (SG). A complete circular capsulorhexis was carried out after instilling methyl cellulose in the anterior chamber and the nucleus aspirated with bimanual irrigation and aspiration technique or by phaco aspiration in calcified thick plate cataracts. All children underwent a primary posterior capsulorhexis through the anterior route with capsulorhexis forceps or vitrector in cases with thick fibrous posterior capsule plaques. A foldable acrylic posterior chamber intra-ocular lens (PCIOL) was implanted in the bag. The scleral tunnel and the side-ports were closed by 10-0 interrupted sutures.

3.6.3.2 *Computational studies*

Computational models and parameters: For the simulations reported in the main body of this manuscript (Figures 3.2 and 3.3), the AlexNet architecture (Krizhevsky et al., 2012) was utilized. The network was implemented in Keras / Tensorflow v2, and training was carried out using the categorical cross-entropy loss function, the Stochastic Gradient Descent (SGD) optimizer with Nesterov momentum of 0.9, a constant learning rate of 0.001, and a batch size of 128. In these simu-

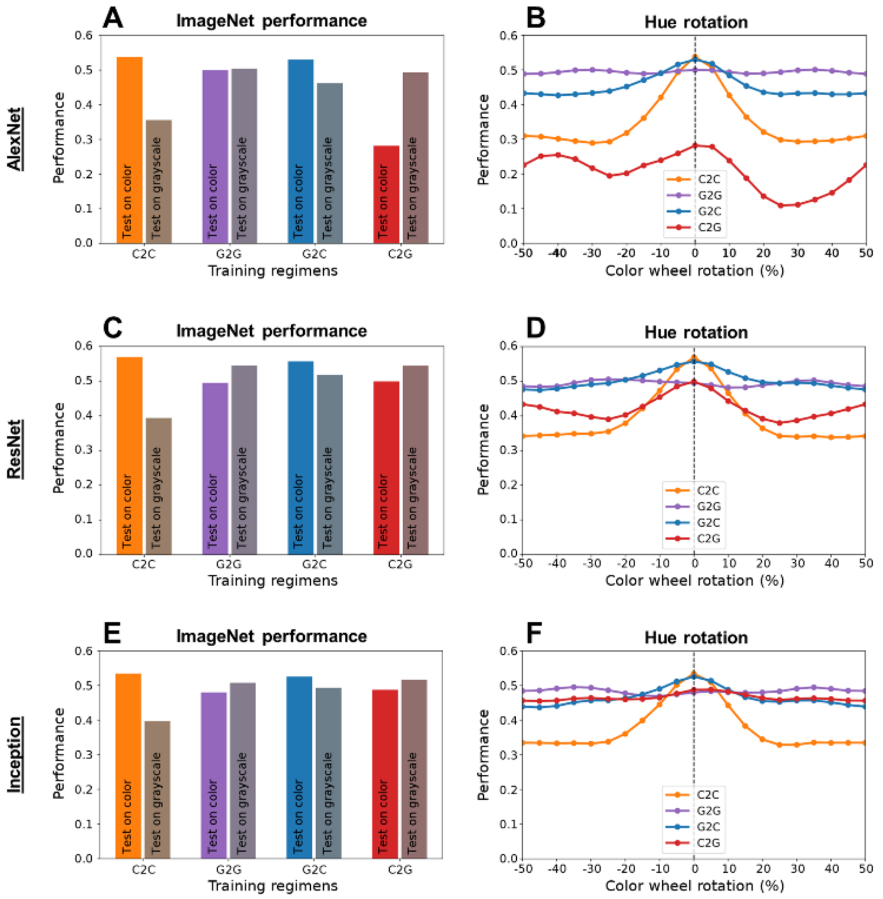


Figure 3.4: (Supplemental Figure) Performance data of the AlexNet (top two panels; based on the same data as Figures 3.2A&B in the main manuscript), ResNet-50 (middle two panels), and Inception v3 network (bottom two panels) when trained using the standard parameter setting from the main manuscript (optimizer = SGD, constant learning rate = 0.001, batch size = 32, Nesterov momentum = 0.9). The training of each regimen lasted for a total of 100 epochs for the AlexNet, 40 epochs for the ResNet, and 20 epochs for the Inception network. The left panels each depict color and grayscale classification performances of networks trained on the ImageNet database using our four different regimens. The right panels each depict classification performances of the four different networks tested on the ImageNet database when rotating the color wheel.

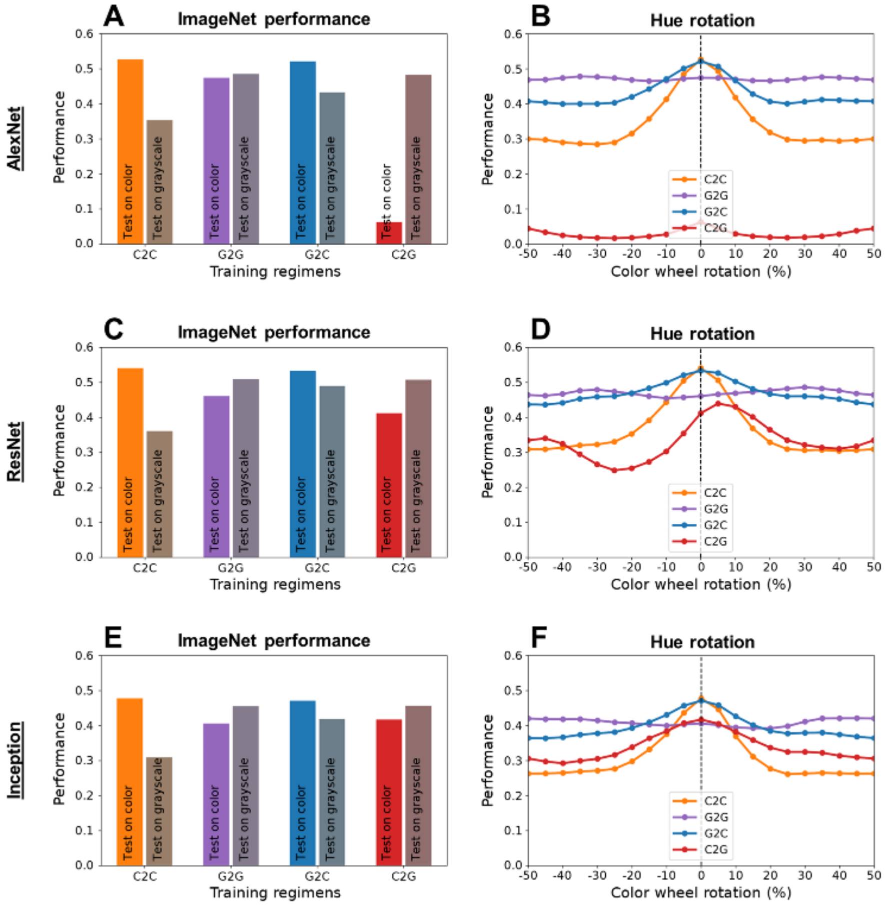


Figure 3.5: (Supplemental Figure) Additional performance data of the AlexNet (top two panels), ResNet-50 (middle two panels), and Inception v3 network (bottom two panels) when trained using a first additional parameter setting (optimizer = SGD, constant learning rate = 0.001, batch size = 64, Nesterov momentum = 0.9, i.e., identical to the standard parameter settings used for Supplemental Figure 3.4, except for a larger batch size). The training of each regimen lasted for a total of 100 epochs for the AlexNet, 40 epochs for the ResNet, and 20 epochs for the Inception network. The left panels each depict color and grayscale classification performances of networks trained on the ImageNet database using our four different regimens. The right panels each depict classification performances of the four different networks tested on the ImageNet database when rotating the color wheel.

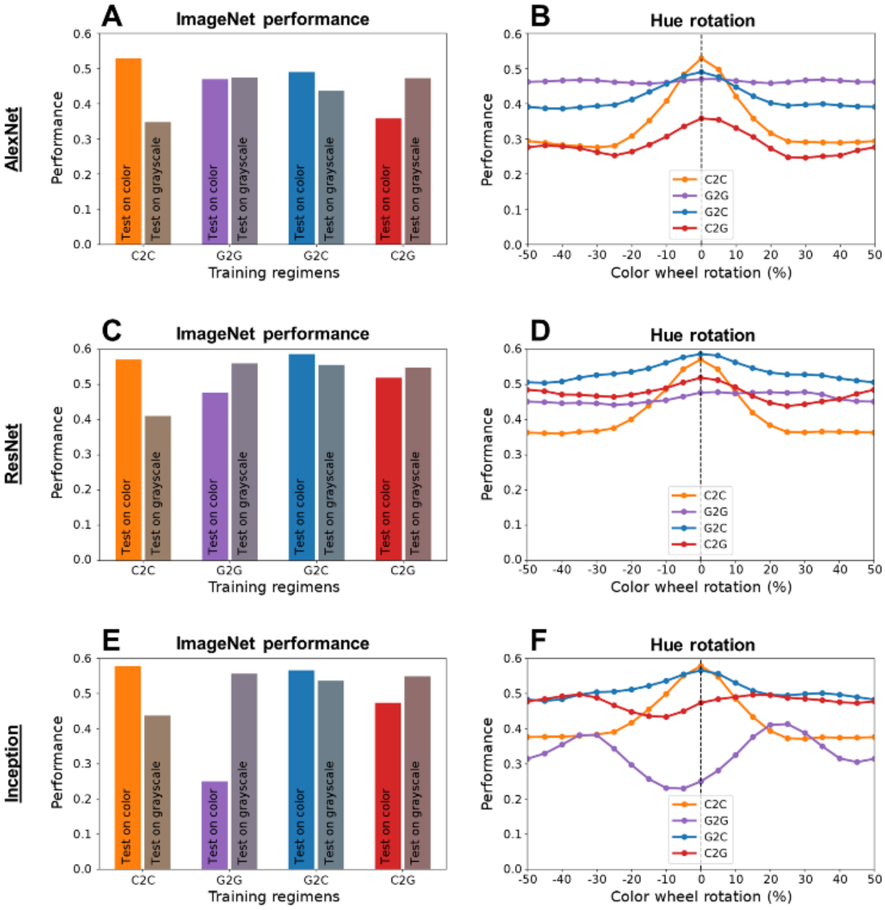


Figure 3.6: (Supplemental Figure) Additional performance data of the AlexNet (top two panels), ResNet-50 (middle two panels), and Inception v3 network (bottom two panels) when trained using a second additional parameter setting (optimizer = Adam, constant learning rate = 0.0001, batch size = 32). The training of each regimen lasted for a total of 100 epochs for the AlexNet, 20 epochs for the ResNet, and 10 epochs for the Inception network. The left panels each depict color and grayscale classification performances of networks trained on the ImageNet database using our four different regimens. The right panels each depict classification performances of the four different networks tested on the ImageNet database when rotating the color wheel.

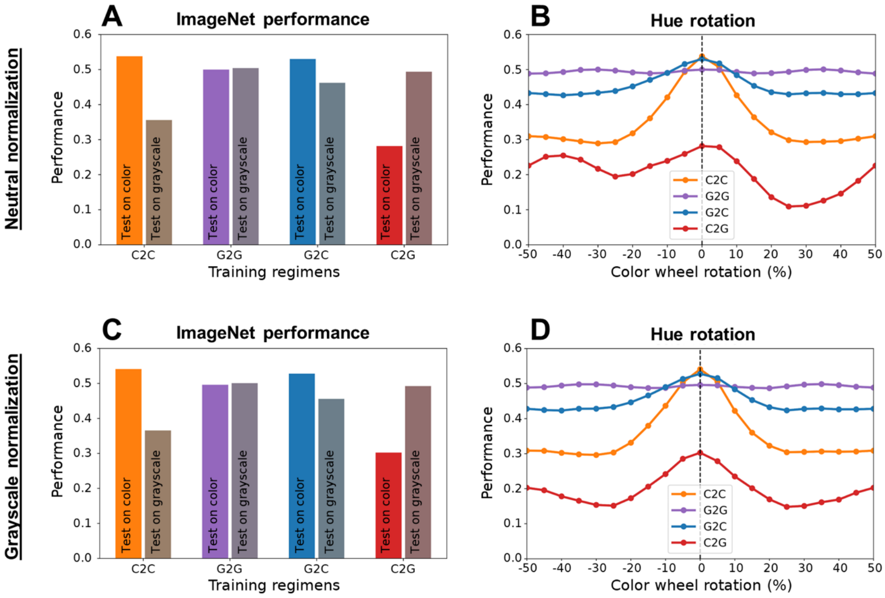


Figure 3.7: (Supplemental Figure) Additional performance data of the AlexNet when trained using the main parameter setting used for Supplemental Figure 3.4 (optimizer = SGD, constant learning rate = 0.001, batch size = 32, Nesterov momentum = 0.9) but using two different choices of normalization: the "neutral" normalization used in the previous figures (simply rescaling pixel values from a [0, 255] to a [-1, 1] range) and an explicit grayscale normalization (shifting and scaling the pixel values into a zero-centered distribution with a standard deviation of 1). The training of each regimen lasted for a total of 100 epochs for the AlexNet, in both scenarios. The left panels each depict color and grayscale classification performances of networks trained on the ImageNet database using our four different regimens. The right panels each depict classification performances of the four different networks tested on the ImageNet database when rotating the color wheel.

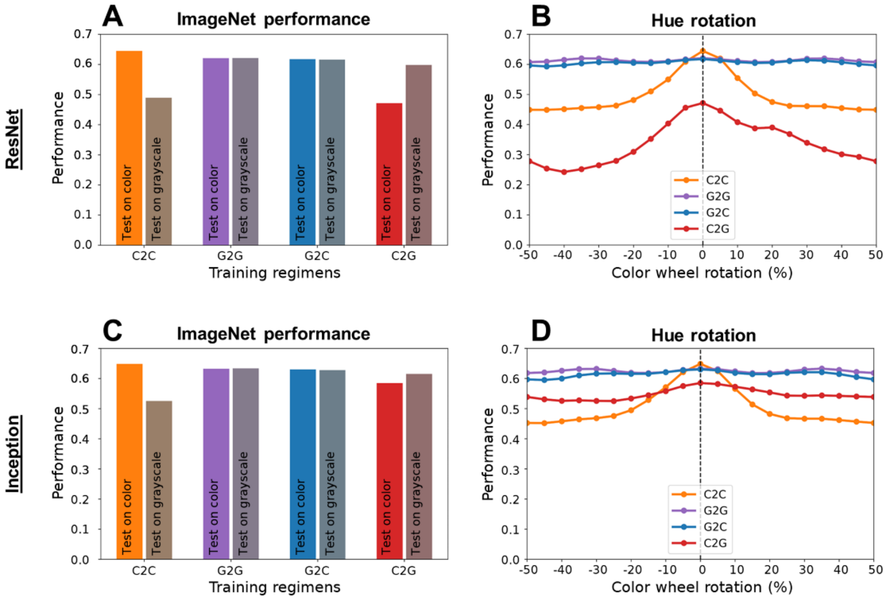


Figure 3.8: (Supplemental Figure) Additional performance data of the ResNet-50 (top two panels) and Inception v3 networks (bottom two panels) when trained using a third additional parameter setting (optimizer = SGD, batch size = 32, Nesterov momentum = 0.9, initial learning rate = 0.1, with a decreasing learning rate schedule monitoring the validation loss (reduction factor = 0.1, patience = 5, minimum delta = 0.0001, cooldown = 0, and minimum learning rate = 0.0001)). The training of each regimen lasted for a total of 40 epochs for the ResNet and 20 epochs for the Inception network. The left panels each depict color and grayscale classification performances of networks trained on the ImageNet database using our four different regimens. The right panels each depict classification performances of the four different networks tested on the ImageNet database when rotating the color wheel.

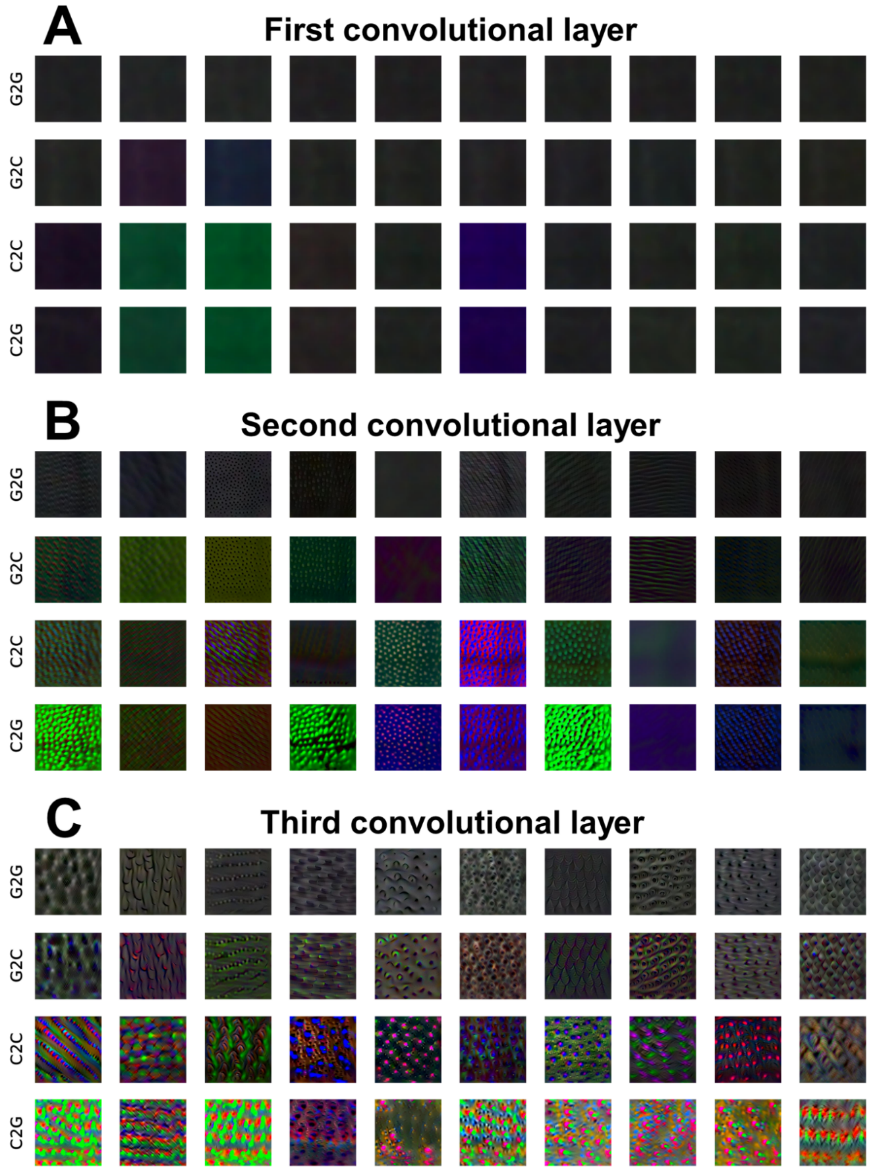

A
First convolutional layer
G2G
G2C
C2C
C2G
B
Second convolutional layer
G2G
G2C
C2C
C2G
C
Third convolutional layer
G2G
G2C
C2C
C2G

Figure 3.9: (Supplemental Figure) Visualization of the synthesized images eliciting maximal activity for the first 10 units of (A) the first convolutional layer, (B) the second convolutional layer, and (C) the third convolution layer, for each of the four models.

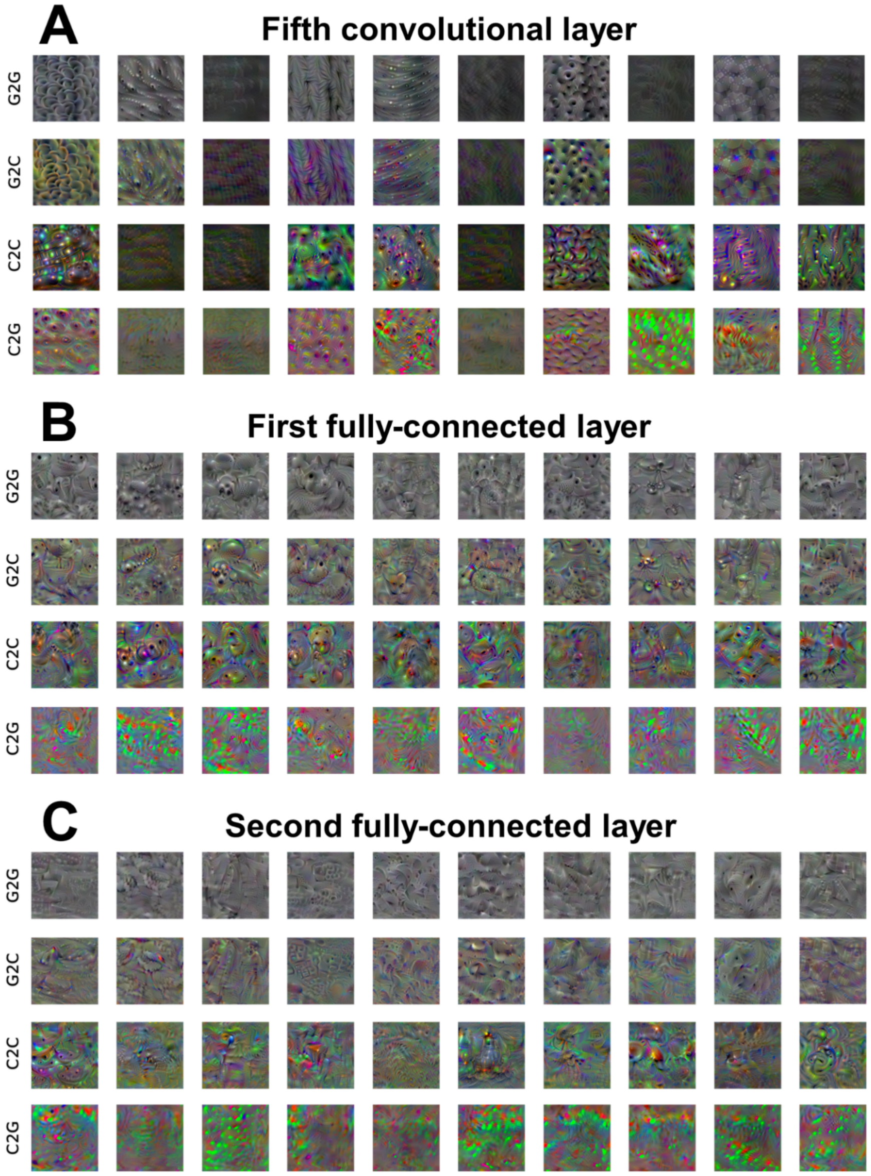

A
Fifth convolutional layer
G2G
G2C
C2C
C2G
B
First fully-connected layer
G2G
G2C
C2C
C2G
C
Second fully-connected layer
G2G
G2C
C2C
C2G

Figure 3.10: (Supplemental Figure) Visualization of the synthesized images eliciting maximal activity for the first 10 units of (A) the fifth convolutional layer, (B) the first fully-connected layer, and (C) the second fully-connected layer (i.e., together with Supplemental Figure 3.9 depicting all layers except for the fourth convolutional and the last fully-connected layer, which are already shown in the main manuscript), for each of the four models.

Subject	Gender	Age at treatment	Blindness detection	Pre-treatment acuity	Post-treatment acuity	Time of testing post-surgery
S1	M	7	At birth	PR/PL	0.802	1.3 years
S2	M	19	At birth	HM	0.89	7 years
S3	M	12	At birth	HM	1.2	11 months
S4	M	15	Within 6 months of birth	FC at 10 cm	0.89	1.2 years
S5	M	9	At birth	PR/PL	1.1	6.1 years
S6	M	16	At birth	HM	0.89	5 years
S7	M	15	At birth	HM	1.1	3 months
S8	M	12	At birth	PR/PL	1.1	3 months
S9	M	9	At birth	PR/PL	1.5	1.3 years
S10	F	11	At birth	PR/PL	1.1	5.9 years

Table 3.1: (Supplemental Table) Patient information of the Prakash individuals who participated in experiment 1, many months or years following their sight-restoring surgeries. PR/PL thereby stands for perception of reflection and light, HM for hand movements, and FC for finger counting.

Table 3.2: (Supplemental Table) Patient information of the Prakash individuals who participated in experiment 2, from a few days post-surgically on. HM thereby stands for hand movements and FC for finger counting

Subject	Gender	Age at treatment	Blindness detection	Pre-op acuity (logMAR)	Post-op1 acuity (logMAR)
S1	F	8	At birth	1.52	1.62
S2	M	20	At birth	2.49	2.3
S3	M	14	At birth	2.44	2.44
S4	F	13	Within 1 year of birth	2.05	2.35
S5	M	12	Within 1 year of birth	1.44	1.41
S6	M	9	Just after the 1st year when he began to walk	2.04	1.82
S7	F	12	Just after the 1st year when she began to walk	1.53	1.52
S8	M	10	At birth	1.42	1.06
S9	M	13	At birth	1.44	1.37
S10	F	12	Within 1 year of birth	2.67	2.36
S11	F	11	Within 1 year of birth	2.67	2.67
S12	F	9	Within 1 year of birth	2.67	2.67
S13	M	8	At 6 months	1.5	1.51
S14	M	9	Within 1 month and confirmed at 2 years	1.80	1.5
S15	F	10	At the age of 2 years	FC close to face	FC at 20 cm
S16	M	10	At birth	HM close to face	5.87
S17	M	14	At birth	1.60	1.50
S18	M	12	At birth	1.70	1.30

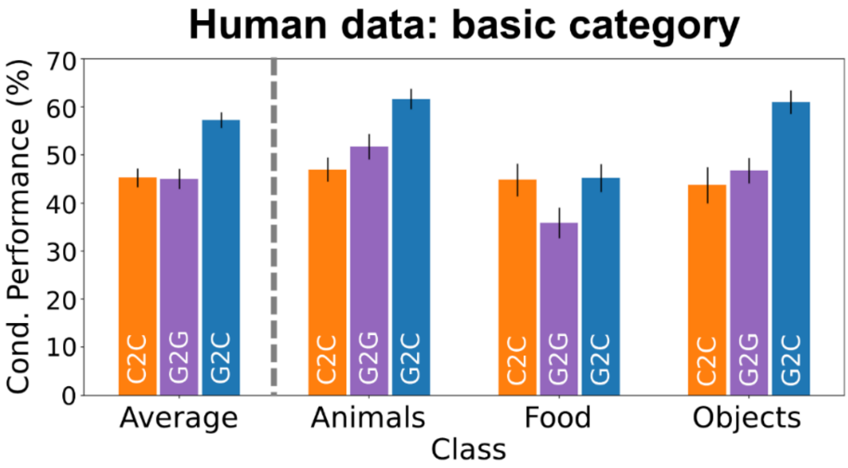


Figure 3.11: (Supplemental Figure) Conditional performance (in % correct) of n = 37 online participants (the data from 2 additional participants could not be used for the conditional computation) in classifying synthesized images into the correct basic-level category, conditioned on their correct classification of the super-category.

Table 3.3: (Supplemental Table) Results of two-way ANOVA comparing performance on the color sensitivity task as a function of group (Prakash group, 20/200 controls, and 20/500 controls) and color channel (R, G, and B).

	Sum of Squares	df	Mean Square	F	p	η^2
Group	27.133	2	13.566	0.083	0.920	0.002
Color	6.304	2	3.152	0.019	0.981	< 0.001
Group * Color	123.361	4	30.840	0.189	0.943	0.010

lations, the learning rate was kept constant, rather than following a decreasing learning rate schedule, in order to ease the interpretation of the number of epochs as well as to ensure that potential benefits in stability of the developmental G2C model, when transitioning from initial grayscale to later full-color training, would not be the mere consequence of a decreased learning rate in the second part of training but would even hold when learning rates are still high later on. The quite standard and simple parameter settings were also chosen to facilitate easy replication. Further, the preprocessing and data augmentation pipelines were kept fairly "vanilla" and included only the simple rescaling of pixel values from a $[0, 255]$ to a $[-1, 1]$ range (henceforth, "neutral" normalization), random horizontal flipping of the images, and, when training on the ImageNet database, the cropping of a random 227 x 227 pixel segment out of the full 256 x 256 pixel images

Table 3.4: (Supplemental Table) Network Configuration Data

	Networks	Total epochs	Optimization	Learning rate (lr)	Batch size	Nesterov momentum	Normalization
Fig. 3.2, 3.3	AlexNet	100	SGD	0.001	128	0.9	"Neutral"
Suppl. Fig. 3.4	AlexNet	100	SGD	0.001	128	0.9	"Neutral"
	ResNet	40	SGD	0.001	128	0.9	"Neutral"
	Inception	20	SGD	0.001	128	0.9	"Neutral"
Suppl. Fig. 3.5	AlexNet	100	SGD	0.001	256	0.9	"Neutral"
	ResNet	40	SGD	0.001	256	0.9	"Neutral"
	Inception	20	SGD	0.001	256	0.9	"Neutral"
Suppl. Fig. 3.6	AlexNet	100	Adam	0.0001	128	-	"Neutral"
	ResNet	20	Adam	0.0001	128	-	"Neutral"
	Inception	10	Adam	0.0001	128	-	"Neutral"
Suppl. Fig. 3.7	AlexNet	100	SGD	0.001	128	0.9	"Normal"
	AlexNet	100	SGD	0.001	128	0.9	"Grayscale"
Suppl. Fig. 3.8	ResNet	40	SGD	lr sched.	128	0.9	"Neutral"
	Inception	20	SGD	lr sched.	128	0.9	"Neutral"

(for the Facescrub database, the input dimensions were adjusted to 100 x 100 pixels, due to the pixel limitations of the database, and no random cropping took place).

To examine the generalizability of the results reported in the main manuscript, we also varied the network architecture (specifically, using the Resnet-50 (He et al., 2016) and Inception v3 (Szegedy et al., 2016) networks) as well as parameter settings (specifically, batch size and learning rate), and explored the impact of decreasing learning rate schedules and normalization choices, as reported in Supplemental Figures 3.4-3.8. Specifically, for the results presented in Supplemental Figure 3.4, we used the same general training settings as in the main body of the manuscript (henceforth, "standard setting") but also trained the ResNet-50 (He et al., 2016) and Inception v3 (Szegedy et al., 2016) architectures on the ImageNet database, each for a different number of overall epochs. For the ResNet-50, the 256 x 256 pixel input images were randomly cropped to 224 x 224 pixel segments as inputs; for the Inception architecture, the 256 x 256 pixel input images were fed directly into the network. For the results presented in Supplemental Figure 3.5, we used all three networks as well as the standard training setting, except for the batch size, which was increased from 32 to 64. For the results presented in Supplemental Figure 3.6, we used all three networks and reduced the batch size to 32 again but utilized the Adam optimizer with a small learning rate of 0.0001. For Supplemental Figure 3.7, we used the standard setting for the AlexNet but varied the choice of normalization, using not only the "neutral" normalization (i.e., simply rescaling pixel values from a $[0, 255]$ range to a $[-1, 1]$ range) but also an explicit grayscale-based normalization (i.e., normalizing based on the mean and variance in the grayscale images, so as to shift and scale the pixel values into a zero-centered distribution with a standard deviation of 1). Both of these choices ensure that for grayscale images, the R, G, and B values are scaled and shifted identically, thereby ensuring that the network does receive effectively achromatic inputs. Finally, for Supplemental Figure 3.8, we utilized a decreasing learning rate schedule for the ResNet-50 and Inception v3 networks (whose performance levels benefit from such schedules especially), monitoring the validation loss, having a reduction factor of 0.1, a patience of 5, a minimum delta of 0.0001, a cooldown of 0, and a minimum learning rate of 0.0001. A direct comparison between the different settings can also be found in Supplemental Table 3.4.

Databases: For all computational figures except Figures 3.2F&G, we utilized the well-established ImageNet object recognition database (Deng et al., 2009), containing more than 1.2 million training images belonging to 1000 different object classes. We used the official split of the database into training and test sets, with the test set containing a total of 50,000 images. To examine the generalizability of our results across databases, and considering that the experience of an infant comprises significant exposure to faces, for the results presented in Figures 3.2F&G of the main body of this manuscript, we also trained the AlexNet on a variant of the FaceScrub face recognition database (Ng and Winkler, 2014), comprising 50429 tightly-cropped 100 x 100 pixel images belonging to 388 different facial identities of celebrities. The dataset was split into 10 parts, and 10-fold cross-validation was carried out for training and testing.

Training regimens: To comprehensively examine the consequences of training on different temporal progressions of chromaticity, we trained different instances of our networks while systematically controlling the availability of color information during training. A total of four different regimens, for each network and setting, were used for training:

- "Color-to-color" (C2C), in which, roughly akin to the experience of Prakash children, the entire training consisted of full-color images,
- "Gray-to-gray" (G2G), in which the entire training consisted of grayscale images,
- "Gray-to-color" (G2C), in which, providing a coarse proxy for the developmental trajectory of typically developing infants, initial training on grayscale images (on half of the total number of epochs) was followed by later training on color images (on half of the total number of epochs), and
- "Color-to-gray" (C2G) in which training followed an inverse-developmental progression but had the same aggregate content as that of the G2C regimen.

For better comparability, for a given network and setting, the total number of epochs (determined to ensure fair convergence of the regimens) were kept identical across the four training regimens. The total number of epochs may, however, differ across networks and parameter settings (see Supplemental Table 3.4 for overview).

Selection of food and fruit classes: To examine network performance on the subset of ImageNet classes containing food and fruit items (as reported in Figures 3.2C&D of the main manuscript), two of the authors of this manuscript independently selected appropriate classes on the following basis: For a class to be classified as a fruit class, it had to meet the botanical definition of fruits. In order for a class to be classified as a food class, it needed to either be a fruit class, depict vegetarian food, or depict non-vegetarian food that can clearly be categorized as food rather than as animal (e.g., burger or meat loaf, but not lobster).

Class color homogeneity index: For the results presented in Figure 3.2E, we extracted color homogeneity indices for each of the 1000

ImageNet classes individually. As a first approximation, we computed the mean hues of all images in the entire ImageNet training set, followed by computing how varied these mean hues are across all instances belonging to each of the 1000 classes. Considering that hues are described in a polar coordinate system, the variation of mean hues was computed, separately for each of the 1000 classes, as follows:

$$\begin{aligned} R &= 1 - \frac{1}{N}\left|\sum_{k=1}^{N} e^{i\theta_k}\right| \\ CCHI &= 1 - R \end{aligned}$$

where θ is the mean hue value in the range $[0, 2\pi)$, N is the number of images in a given class, and R is termed the mean resultant length (see, for instance, Fisher, 1995), which ranges from 0 to 1, with 1 indicating no variation in mean hues across the different images belonging to a given class and 0 representing maximal variation. As we wished to extract a class color homogeneity index (CCHI), where low values indicate low variation and high values indicate high variation, we defined it as the opposite of R (i.e., $1 - R$).

First-layer receptive field analysis: For the results presented in Figure 3.3B of the main manuscript, in order to quantify the presence of color in a given receptive field, we first defined a measure to capture the imbalance of intensities in the three color channels for each individual pixel:

$$\begin{aligned} x &= R\cos(0) + G\cos\left(\frac{2}{3}\pi\right) + B\cos\left(-\frac{2}{3}\pi\right) \\ y &= R\sin(0) + G\sin\left(\frac{2}{3}\pi\right) + B\sin\left(-\frac{2}{3}\pi\right) \\ m &= \sqrt{x^2 + y^2} \end{aligned}$$

where R, G, and B represent the respective channel intensities at a given pixel. We thereby chose a single-pixel measure of color channel imbalances, rather than measures necessitating spatial averaging, to avoid cancelling out opponent colors. Finally, we summarized the 11 x 11 individual pixel values of colorfulness, for each receptive field, into a single score by computing the average.

Synthesizing deeper-layer stimuli eliciting maximal activation: To examine deeper-layer representations, utilizing the methodology of Olah et al. (2017), we synthesized images eliciting maximal activation for each of the different units in each of the different layers for each of the four trained networks. We thereby used the default settings of the software package available at https://github.com/keisen/tf-keras-vis and chose a total of 250 steps of optimization for extracting the stimuli.

3.6.3.3 *Online studies*

Class selection: To examine the recognizability of the synthesized images for the final classification layer of the Alexnet, when trained with different regimens, we designed a perceptual online study. As we also wished to probe whether findings related to recognizability would differ between ecologically-diverse higher-level categories, we decided to pick classes mapping onto three super-categories: food, animals, and objects. We did so with the following systematicity:

As there were only relatively few food classes and as we were seeking to balance the number of basic-level classes belonging to the three bigger categories, we began by making a sub-selection of the food classes that we had determined earlier (see section "Selection of food and fruit classes"). Specifically, we selected all previously identified food classes except the ones that two independently-judging authors of this manuscript were not able to name explicitly (e.g., custard apple), judged as being too similar (e.g., bread and dough), or those that could lead to confusion in the super-category classification (e.g., hotpot could be seen as food or as a pot, which would be part of the super-category "objects"). Overall, only few classes were disregarded, yielding a total of 32 food classes.

We next sought to find 32 basic-level classes belonging to the super-category of animals as well as 32 basic-level classes belonging to the super-category of objects – a small subset of the many more object and animal classes that are part of the ImageNet categories. We thereby wished to avoid a purely subjective selection and also wanted to ensure that different basic-level categories can be named and distinguished properly (e.g., not featuring 32 different dog types). We therefore proceeded by utilizing the categories reported in the Supporting Information of Mehrer et al. (2021), in which 565 basic-level categories are listed, sorted on the basis of human concreteness ratings and linguistic usage statistics, in the researchers' attempt to reveal the most relevant categories for humans. Two authors of the present manuscript went through this list and independently identified which basic-level classes clearly belonged to the animal or the object super-categories. Taking into account all classes for which the two authors agreed, they then iterated through the list, from top to bottom, and picked the first 32 basic-level classes for each of the two remaining broader categories, assuming they mapped clearly onto a class present in the ImageNet database (note that if the basic-level class mapped onto several subordinate classes, we chose a representative subordinate class). Together, this led to three sets of 32 image classes each belonging to the 3 super-categories (food, animals, and objects).

Experimental procedure: Before the start of the experiment, participants were informed that they would see versions of images resembling certain animals, foods, or other objects. They were instructed that, in each trial, they would see an image and would need to click with their mouse on one of the labels "animal", "food", or "object". They were informed that after this initial choice, the image would disappear and that they would need to select one out of the 32 different object labels (within the previously selected super-category) presented on the screen (e.g., "cat", "dog", "tiger", and 29 others if their initial choice was "animal"). They were also instructed to guess in case they did not know the answer. There were no time restrictions for completing a given trial. Note that participants were informed that the class "food" contains either vegetarian food or meat/fish as part of a prepared meal (e.g., in a burger or meat loaf); if they saw an entire lobster or fish, they were informed that the desired category would be "animal", not "food". They were also informed that the class "objects" does not contain items related to animals or food.

In order to reduce the overall experiment time for each individual participant, the 96 extracted classes were split into two non-overlapping sets A and B, each of which contained 16 classes belonging to each of the three higher-level categories (food, animals, or objects). For both sets of 48 classes, we extracted the synthesized stimuli eliciting maximal activity for each of the G2G, G2C, and C2C networks (the C2G stimuli were not used as they appeared entirely unrecognizable), resulting in a total of 144 images to be presented. These images were presented in random order (randomized differently across participants). In addition, following the 144 images, 48 representative real-world images of the categories were shown, to examine whether a given participant was familiar with the image classes presented. Taken together, our experiment had 192 trials and took approximately 30 minutes until completion.

Running the online study: The experiment was implemented using Psychopy, run on Pavlovia, and participants were recruited using Prolific. To be eligible for this study, participants were required to have normal or corrected-to-normal acuity, no color blindness, and be fluent in English. A total of 40 participants were recruited, and the data from all except one participant (who performed poorly in the control task at the end of the experiment) were used for the analysis reported in the main manuscript. 21 of the 39 valid online participants belonged to the set A split, and 18 participants belonged to the set B split. For the results reported in Supplemental Figure 3.11 (performance on basic-level classification conditioned on correct superclass classification), two additional subjects had to be excluded from the analysis as the conditional probability could not be computed (for at least one super-category and model combination, they had never gotten the super-category correct).

REFERENCES

Biederman, Irving and Ginny Ju (1988). "Surface versus edge-based determinants of visual recognition." In: *Cognitive psychology* 20.1, pp. 38–64.

Cadena, Santiago A, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, and Alexander S Ecker (2019). "Deep convolutional models improve predictions of macaque V1 responses to natural images." In: *PLoS computational biology* 15.4, e1006897.

Carvalho, Livia S, Daniel MA Pessoa, Jessica K Mountford, Wayne IL Davies, and David M Hunt (2017). "The genetic and evolutionary drives behind primate color vision." In: *Frontiers in ecology and Evolution* 5, p. 34.

Davidoff, JB and AL Ostergaard (1988). "The role of colour in categorial judgements." In: *The Quarterly Journal of Experimental Psychology Section A* 40.3, pp. 533–544.

DeLoache, Judy S, Mark S Strauss, and Jane Maynard (1979). "Picture perception in infancy." In: *Infant behavior and development* 2, pp. 77–89.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei (2009). "Imagenet: A large-scale hierarchical image database." In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Deregowski, Jan B, Elizabeth S Muldrow, and WF4680942 Muldrow (1972). "Pictorial recognition in a remote Ethiopian population." In: *Perception* 1.4, pp. 417–425.

Dominy, Nathaniel J and Peter W Lucas (2001). "Ecological importance of trichromatic vision to primates." In: *Nature* 410.6826, pp. 363–366.

Elder, James H and Ljiljana Velisavljević (2009). "Cue dynamics underlying rapid detection of animals in natural scenes." In: *Journal of Vision* 9.7, pp. 7–7.

Fisher, Nicholas I (1995). *Statistical analysis of circular data*. cambridge university press.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016). "Deep residual learning for image recognition." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hiramatsu, Chihiro, Amanda D Melin, William L Allen, Constance Dubuc, and James P Higham (2017). "Experimental evidence that primate trichromacy is well suited for detecting primate social colour signals." In: *Proceedings of the Royal Society B: Biological Sciences* 284.1856, p. 20162458.

Hochberg, Julian and Virginia Brooks (1962). "Pictorial recognition as an unlearned ability: A study of one child's performance." In: *the american Journal of Psychology* 75.4, pp. 624–628.

Hurlbert, Anya (2003). "Colour vision: primary visual cortex shows its influence." In: *Current Biology* 13.7, R270–R272.

Jacobs, Gerald H (2008). "Primate color vision: a comparative perspective." In: *Visual neuroscience* 25.5-6, pp. 619–633.

Jacobs, Rachel L, Carrie C Veilleux, Edward E Louis, James P Herrera, Chihiro Hiramatsu, David C Frankel, Mitchell T Irwin, Amanda D Melin, and Brenda J Bradley (2019). "Less is more: lemurs (Eulemur spp.) may benefit from loss of trichromatic vision." In: *Behavioral Ecology and Sociobiology* 73, pp. 1–17.

Jones, Alex L (2018). "The influence of shape and colour cue classes on facial health perception." In: *Evolution and Human Behavior* 39.1, pp. 19–29.

Kawamura, Shoji (2016). "Color vision diversity and significance in primates inferred from genetic and field studies." In: *Genes & genomics* 38.9, pp. 779–791.

Kozlovskiy, Stanislav and Anton Rogachev (2021). "How Areas of Ventral Visual Stream Interact When We Memorize Color and Shape Information." In: *Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics: Proceedings of the 9th International Conference on Cognitive Sciences, Intercognsci-2020, October 10-16, 2020, Moscow, Russia 9*. Springer, pp. 95–100.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems* 25.

Lindsay, Grace W (2021). "Convolutional neural networks as a model of the visual system: Past, present, and future." In: *Journal of cognitive neuroscience* 33.10, pp. 2017–2031.

Livingstone, Margaret and David Hubel (1988). "Segregation of form, color, movement, and depth: anatomy, physiology, and perception." In: *Science* 240.4853, pp. 740–749.

Mapelli, Daniela and Marlene Behrmann (1997). "The role of color in object recognition: Evidence from visual agnosia." In: *Neurocase* 3.4, pp. 237–247.

Marx, Svenja, Onno Hansen-Goos, Michael Thrun, and Wolfgang Einhäuser (2014). "Rapid serial processing of natural scenes: color modulates detection but neither recognition nor the attentional blink." In: *Journal of Vision* 14.14, pp. 4–4.

Mehrer, Johannes, Courtney J Spoerer, Emer C Jones, Nikolaus Kriegeskorte, and Tim C Kietzmann (2021). "An ecologically motivated image dataset for deep learning yields better models of human vision." In: *Proceedings of the National Academy of Sciences* 118.8, e2011417118.

Melin, Amanda D, Omer Nevo, Mika Shirasu, Rachel E Williamson, Eva C Garrett, Mizuki Endo, Kodama Sakurai, Yuka Matsushita, Kazushige Touhara, and Shoji Kawamura (2019). "Fruit scent and observer colour vision shape food-selection strategies in wild capuchin monkeys." In: *Nature communications* 10.1, p. 2407.

Meng, Ming and Mary C Potter (2008). "Detecting and remembering pictures with and without visual noise." In: *Journal of Vision* 8.9, pp. 7–7.

Nestor, Adrian and Michael J Tarr (2008). "Gender recognition of human faces using color." In: *Psychological Science* 19.12, pp. 1242–1246.

Nevo, Omer, Kim Valenta, Diary Razafimandimby, Amanda D Melin, Manfred Ayasse, and Colin A Chapman (2018). "Frugivores and the evolution of fruit colour." In: *Biology Letters* 14.9, p. 20180377.

Ng, Hong-Wei and Stefan Winkler (2014). "A data-driven approach to cleaning large face datasets." In: *2014 IEEE international conference on image processing (ICIP)*. IEEE, pp. 343–347.

Odell, Naomi V, David A Leske, Sarah R Hatt, Wendy E Adams, and Jonathan M Holmes (2008). "The effect of Bangerter filters on optotype acuity, Vernier acuity, and contrast sensitivity." In: *Journal of American Association for Pediatric Ophthalmology and Strabismus* 12.6, pp. 555–559.

Olah, Chris, Alexander Mordvintsev, and Ludwig Schubert (2017). "Feature visualization." In: *Distill* 2.11, e7.

Regan, Benedict C, C Julliot, B Simmen, Françoise Viénot, P Charles-Dominique, and John D Mollon (2001). "Fruits, foliage and the evolution of primate colour vision." In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356.1407, pp. 229–283.

Righi, Giulia and Jean Vettel (2011). "Ventral Visual Pathway." In: *Encyclopedia of Clinical Neuropsychology*. Ed. by Jeffrey S. Kreutzer, John DeLuca, and Bruce Caplan. New York, NY: Springer New York, pp. 2598–2600. ISBN: 978-0-387-79948-3.

Schrimpf, Martin, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo (2020). "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?" In: *bioRxiv*.

Shapley, Robert (1990). "Visual sensitivity and parallel retinocortical channels." In: *Annual review of psychology* 41.1, pp. 635–658.

Shapley, Robert and Michael J Hawken (2011). "Color in the cortex: single-and double-opponent cells." In: *Vision research* 51.7, pp. 701–717.

Shinskey, Jeanne L and Liza J Jachens (2014). "Picturing objects in infancy." In: *Child development* 85.5, pp. 1813–1820.

Sinha, Pawan (2013). "Once blind and now they see." In: *Scientific American* 309.1, pp. 48–55.

Skelton, Alice E, John Maule, and Anna Franklin (2022). "Infant color perception: Insight into perceptual development." In: *Child Development Perspectives* 16.2, pp. 90–95.

Spence, Charles (2018). "Background colour & its impact on food perception & behaviour." In: *Food Quality and Preference* 68, pp. 156–166.

Stephen, Ian D, Miriam J Law Smith, Michael R Stirrat, and David I Perrett (2009). "Facial skin coloration affects perceived health of human faces." In: *International journal of primatology* 30, pp. 845–857.

Storrs, Katherine R, Tim C Kietzmann, Alexander Walther, Johannes Mehrer, and Nikolaus Kriegeskorte (2021). "Diverse deep neural networks all predict human inferior temporal cortex well, after training and fitting." In: *Journal of cognitive neuroscience* 33.10, pp. 2044–2064.

Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna (2016). "Rethinking the inception architecture for computer vision." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Teller, Davida Y (1998). "Spatial and temporal aspects of infant color vision." In: *Vision Research* 38.21, pp. 3275–3282.

Thorstenson, Christopher A, Adam D Pazda, and Andrew J Elliot (2020). "Social perception of facial color appearance for human trichromatic versus dichromatic color vision." In: *Personality and Social Psychology Bulletin* 46.1, pp. 51–63.

Tusa, Ronald J, MX Repka, Carolyn B Smith, and SJ Herdman (1991). "Early visual deprivation results in persistent strabismus and nystagmus in monkeys." In: *Investigative ophthalmology & visual science* 32.1, pp. 134–141.

Vogelsang, Lukas, Sharon Gilad-Gutnick, Evan Ehrenberg, Albert Yonas, Sidney Diamond, Richard Held, and Pawan Sinha (2018). "Potential downside of high initial visual acuity." In: *Proceedings of the National Academy of Sciences* 115.44, pp. 11333–11338.

Vogelsang, Marin, Lukas Vogelsang, Sidney Diamond, and Pawan Sinha (2023). "Prenatal auditory experience and its sequelae." In: *Developmental Science* 26.1, e13278.

Yuodelis, Cristine and Anita Hendrickson (1986). "A qualitative and quantitative analysis of the human fovea during development." In: *Vision research* 26.6, pp. 847–855.

4

ON THE ORIGIN OF THE PARVO- AND MAGNOCELLULAR DIVISION: POTENTIAL ROLE OF DEVELOPMENTAL EXPERIENCE

Content from

Vogelsang, M., Vogelsang, L., Pipa, G., Diamond, S., & Sinha, P. (Submitted). "On the origin of the parvo- and magnocellular division: potential role of developmental experience". **Submitted manuscript** (initial submission: October 2023).

4.1 ABSTRACT

Neurons in the mammalian early visual pathway can be broadly segregated into two groups: magnocellular - exhibiting low spatial frequency and low chromatic sensitivity, and parvocellular - exhibiting high spatial frequency and high chromatic sensitivity. While this division is widely acknowledged, its genesis remains unclear. We here propose an account based on early developmental trajectories of sensory experience. Specifically, we hypothesize that the temporal confluence of spatial frequency and chromatic sensitivities during early development may shape neuronal response properties characteristic of this division. Results of computational simulations with deep neural networks trained on developmentally-inspired 'biomimetic' protocols provide strong support for this hypothesis. Further, biomimetic training induced a more human-like global shape bias in classification, driven by receptive fields exhibiting magnocellular characteristics. Together, these results provide a potential account for the origin of a prominent organizing principle in the mammalian visual system and point to improved deep network training procedures.

4.2 INTRODUCTION

Cells in the early visual pathway in mammalian brains can be broadly segregated into two groups: magnocellular and parvocellular (reviewed in Livingstone and Hubel, 1988; Shapley, 1990). Two key characteristics of this division, already strongly evident by the lateral geniculate nucleus and originating at the level of retinal ganglion cells (Shapley, 1992), are related to color (Hicks et al., 1983; Hubel and Livingstone, 1990; Schiller and Malpeli, 1978; Wiesel and Hubel, 1966) and spatial frequency sensitivities (Derrington and Lennie, 1984; O'Keefe et al., 1998; Usrey and Reid, 2000). Magno units exhibit receptive fields (rfs) markedly larger than those of parvo cells and are mostly achromatic. In contrast, most parvo cells have small rfs and are strongly tuned to color content. Thus, the magnocellular group exhibits both low spatial frequency and chromatic sensitivity, and the parvocellular group exhibits high spatial frequency and chromatic tuning. In concert also with the divisions' temporal characteristics, notably the magno cells' faster conductance (Dreher et al., 1976; Schiller and Malpeli, 1978) and higher temporal frequency sensitivity (Derrington and Lennie, 1984; O'Keefe et al., 1998; Usrey and Reid, 2000), the magno pathway has been implicated in fast, course-grained spatial analyses, whereas the parvo pathway is believed to be responsible for the analysis of fine-grained spatial and chromatic information. While the anatomical and physiological division between the two pathways is widely accepted and has also been complemented by psychophysical and clinical studies (e.g., Livingstone and Hubel, 1987; Livingstone et al., 1991), its genesis is not yet clearly established. Here, we propose and computationally test an account of the origin of the magno-parvo distinction, which is based on early developmental trajectories of sensory experience.

As is the case with many dimensions of human perceptual development, color sensitivity (Adams and Courage, 2002; Dobkins et al., 1997) and visual acuity (Courage and Adams, 1990; Dobson and Teller, 1978) mature over the months following birth from limited to proficient. The underlying factors are believed to be the maturation of retinal photoreceptor morphology and transductional efficiency, as well as elaboration of circuits in the retina and cortex (Banks and Bennett, 1988; Candy and Banks, 1999; Jacobs and Blakemore, 1988; Kiorpes and Movshon, 2004). The joint developmental progression of these two perceptual dimensions could causally influence how they are encoded in the neural substrates. Specifically, the start of visual experience is accompanied by low acuity and poor color sensitivity. Hence, the cell response properties that emerge at this time could come to jointly encode these two attributes. However, as development progresses, higher acuity and richer color information become available and may be conjoined in neuronal response properties. Thus, the joint coding of low spatial frequency and low color information in some units, and high spatial frequency and high color sensitivity in others, could be an outcome of the co-occurrence of these features at different time points during development.

Here, we tested this account through simulations with deep neural networks as computational models. While not without their limitations as models of the biological system, they have proven useful for predicting human performance and neural activities (e.g., Lindsay, 2021; Schrimpf et al., 2020). Of particular relevance to us, they provide a systematic methodology for directly probing the consequences of deliberate manipulations of sensory experience, which behavioral studies with human participants do not allow due to obvious ethical reasons. Specifically, one can expose a deep network, whose internal representations are learned based on the stimuli to which it is exposed, to different temporal progressions of inputs and subsequently study the consequences of training. Some of these progressions are thereby chosen to recapitulate aspects of typical development, while others serve as non-developmental controls. In the specific context of this study, this allows us to examine the differently trained networks' receptive fields in the early stages of their architecture in terms of their color and spatial frequency tuning; it also permits us to probe the functional roles of these receptive fields for the network behavior in ecologically-relevant contexts.

4.3 methods

We utilized the AlexNet architecture (Krizhevsky et al., 2012) and trained it on the ImageNet object database (Deng et al., 2009). To avoid artificially restricting the kinds of receptive field structures that can be learned, and to allow for more precise frequency-based analyses, we increased the permissible size of the receptive fields in the first convolutional layer from 11x11 to 22x22 pixels. Furthermore, to prevent the emergence of units primarily coding for noise, we reduced the overall number of receptive fields in the first layer from 96 to 48 (however, results of the simulations carried out with 96 receptive fields, revealing qualitatively similar findings, are provided as Supplemental Figure 4.4). We trained this network on two temporal stimulus progressions, or 'training regimens', of interest ('standard' and 'biomimetic'):

i In the 'standard' regimen, as a non-developmental control, we trained our network on high-resolution, full-color images for the entire training duration comprising 200 epochs. This is the typical procedure followed in deep convolutional network training (e.g., Goodfellow et al., 2016).

ii In the developmentally-based 'biomimetic' regimen, we trained the network on reduced resolution, achromatic images for the first 100 epochs, and on high-resolution, full-color images for the subsequent 100 epochs.

For completeness, and to examine the generalizability of our findings, we also trained the network on three additional developmentally-inspired regimens, with results provided in Supplemental Figures 4.5 & 4.6:

iii Given the faster development of color than acuity, resulting in an intermediate stage where color sensitivity is fully developed but visual acuity is not (Adams and Courage, 2002; Courage and

Adams, 1990), in the 'biomimetic v2' regimen, we split the first 100 epochs of degraded training into 50 epochs of low-resolution, achromatic and 50 epochs of low-resolution, full-color training.

iv In the 'biomimetic v3' regimen, we extended the second phase of training from 100 to 200 epochs in light of the relatively short duration of initially degraded visual experience in human development (Adams and Courage, 2002; Courage and Adams, 1990; Dobkins et al., 1997; Dobson and Teller, 1978).

v In the 'biomimetic v4' regimen, we restricted a proportion (here, 50%) of the first-layer filters to only learn during the second half of training. This serves as an additional biomimetic control in light of evidence that parvocellular cells appear to develop after magnocellular ones (Rakic, 1977) and may thus not have effectively been exposed to maximally degraded stimuli.

Further methodological details of training and analysis can be found in the Supplemental Information, along with a visualization of the different training regimens (see Supplemental Figure 4.7).

4.4 RESULTS

4.4.1 *Learned receptive field structures*

Figures 4.1A&B depict the learned receptive fields in the first convolutional layer of our network following training with the 'standard' regimen on the one hand and a developmentally-inspired 'biomimetic' regimen on the other. As is evident by visual inspection, and as quantified in Figures 4.1C&D showing the distribution of individual receptive fields' chromatic and spatial frequency tuning, training with the biomimetic regimen results in receptive fields that are tuned significantly less to high spatial frequency and chromatic content ($t(94) = 2.75$, $p = 0.007$ and $t(94) = 3.43$, $p < 0.001$ in two-tailed two-sample t-tests comparing the color and spatial frequency indices of the two models, respectively). This effect highlights the significance of spatially-extended and luminance-based receptive fields instantiated during the first half of training. Notably, these differences in rf sensitivity distributions persist notwithstanding the second half of the training with high resolution and high chromatic content stimuli.

Of greatest relevance to the magno-parvo distinction, Figures 4.1E&F depict the joint distribution of frequency and color coding of individual receptive fields, revealing marked differences between results of the standard and biomimetic training regimens. In the 'standard' network, no clear relationship between the two attributes is evident, except for the existence of a few high-frequency achromatic receptive fields (see Figure 4.1E, blue ellipse). These, however, do not map onto

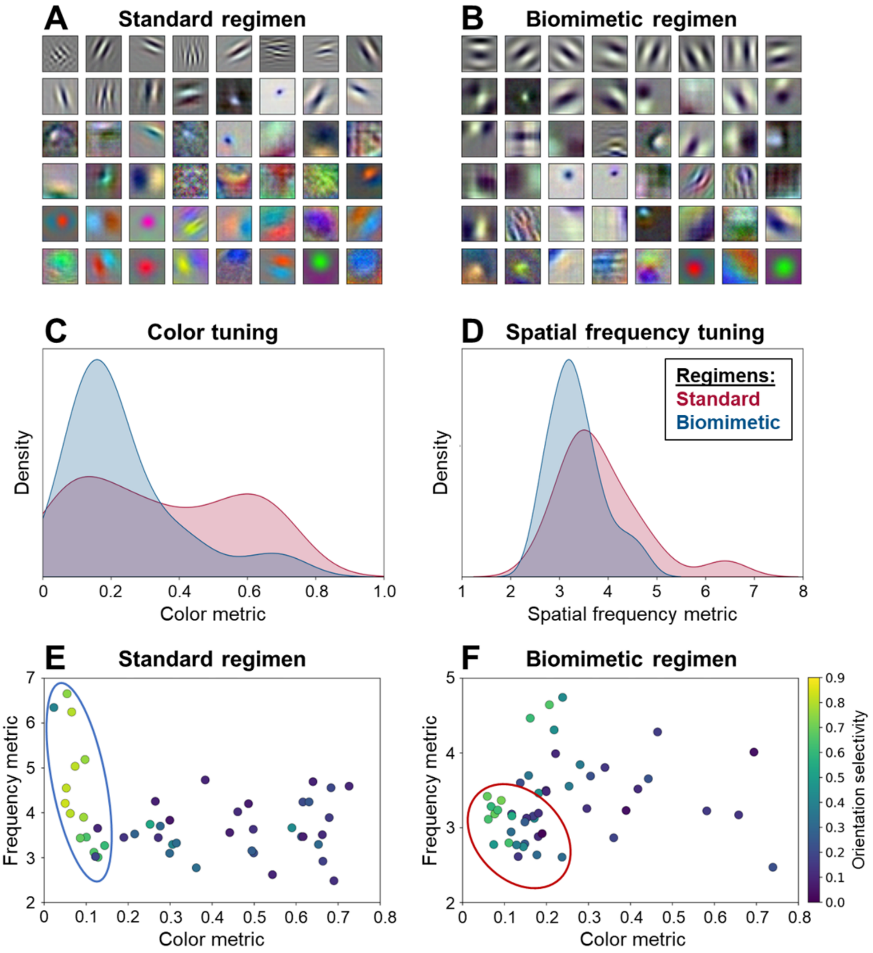


Figure 4.1: **A&B.** First-layer receptive fields following training with the standard (A) and biomimetic (B) regimen. **C&D.** Color (C) and spatial frequency (D) distribution of individual receptive fields. **E&F.** Scatter plot depicting the joint frequency and color coding of individual receptive fields following training with the standard (E) and biomimetic regimen (F).

either the magnocellular or the parvocellular group of neural units but rather reveal an anti-correlation between the two attributes. In the biomimetic network, by contrast, we observe a clear cluster of magnocellular-like receptive fields tuned to both low frequency and low chromatic content (see Figure 4.1F, red ellipse). While the units of the biomimetic model tuned to high spatial frequency or high chromatic content exhibit a greater heterogeneity than the magnocellular-like ones (see Figure 4.1F, dots outside of the red ellipse), this observation is not unexpected. The well-separated parvo and magno layers in the LGN have been reported to mix in the visual cortex (Hubel and Livingstone, 1990). This results in two subpopulations of parvocellular units, both of which are not entirely color-blind: interblobs, exhibiting high spatial frequency selectivity and strong orientation tuning on the one hand, and blobs exhibiting lower spatial frequency selectivity and low orientation tuning on the other (Hubel and Livingstone, 1990). A closer examination of the individual data points in Figure 4.1F, where colors code for a given unit's orientation selectivity, reveals that the units outside of the highlighted magnocellular cluster that are tuned to higher frequencies have a slight tendency to indeed be more orientation-selective than those with lower spatial frequency tuning.

Taken together, we conclude that training with the biomimetic regimen results in the emergence of a relatively homogeneous magnocellular group of units, which is markedly absent in the non-developmentally trained network, as well as different cell types that are more aligned with parvocellular characteristics. These general patterns also hold for the other biomimetic control regimens that were tested (see Supplemental Figures 4.4-4.6).

4.4.2 *Texture/shape-bias in classification decisions*

Next, considering that the magno pathway is believed to be responsible for more coarse-grained processing while the parvo pathway is specialized for fine-grained spatial analysis, we examined the relationship between the receptive fields and the network's performance. An important dimension in this regard is that of local texture versus global shape encoding. We tested whether classification decisions are biased toward texture or shape, using the texture-shape conflict methodology detailed in Geirhos et al. (2018). First, as can be seen in Figures 4.2A&B, the biomimetic model exhibits a markedly stronger bias to classify images based on shape, indicating its classification strategy to be more similar to that of humans (Geirhos et al., 2018). In contrast, the standard training did not lead to any such bias.

Further, we sought to determine whether it is the population of units exhibiting magnocellular characteristics that causally induces the stronger shape bias of the biomimetic model. To this end, we gradually eliminated (or 'ablated') the most color-tuned vs. the least color-tuned

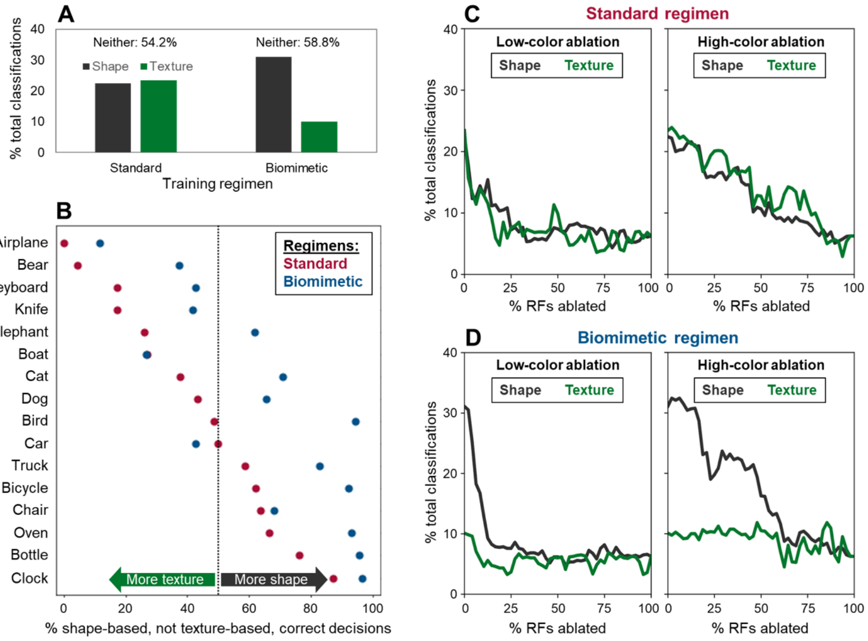


Figure 4.2: **A.** Percentage of total classifications correct in terms of shape, correct in terms of texture, or neither. **B.** Percentage of shape-based, as opposed to texture-based, correct classifications, for each of the 16 different super-classes used. **C&D.** Shape-texture bias as a function of the number of the ablated least colorful (left) and most colorful (right) units, for the standard (C) and biomimetic (D) regimen.

receptive fields of the trained networks and re-computed the shape bias (we chose low color, rather than low frequency, as a proxy index for magnocellular units considering the greater homogeneity observed in Figure 4.1). This analysis, depicted in Figures 4.2C&D, reveals that ablating less than 25% of the least colorful receptive fields entirely eliminates the shape bias of the biomimetically trained network. By contrast, eliminating the same number of the most color-tuned receptive fields does not have a comparable impact. For the 'standard' network, we do not observe any differential effect between the two forms of ablation. To sum, in accordance with what would be expected to be the psychophysical correlates of this division, training with our biomimetic regimen induced not only more human-like classification decisions based on global shape rather than local texture information but also revealed the causal role of receptive fields exhibiting magnocellular characteristics in supporting such global shape bias. These findings also generalize across the other biomimetic regimens tested (see Supplemental Figure 4.5).

4.4.3 *Unit ablation and invariance studies*

To complement the above results on shape versus texture bias in encoding, we also examined the effects of ablation on the classification performance of both models (see Figures 4.3A&B). Similar to the differences reported in Figure 4.2C, the biomimetic model is more differentially affected than the standard one by the ablation of low color or low spatial frequency units, relative to the ablation of high color or high spatial frequency units. Further, the relative importance of these units remains more similar across full-color vs. grayscale images for the biomimetic network than for the standard network, presumably due to increased invariance to the removal of chromatic information.

To quantitatively examine the invariance of both networks to the removal of chromatic and high spatial frequency content, we determined the distribution of all units' correlation coefficients when presented with color vs. grayscale (see Figure 4.3C) and full-frequency vs. blurred (see Figure 4.3D) images. This analysis reveals markedly greater invariance of the biomimetic model for both stimulus dimensions across almost all network layers. In addition, we examined the invariance to the reduction of image contrast, as reported in Figure 4.3E. While this analysis did not reveal apparent differences between the two networks, we observe a notable pattern upon ablation: when ablating the least colorful filters in the first layer, the biomimetic model becomes less invariant to contrast reduction than the standard model, but when ablating the most colorful units, it becomes relatively more invariant (see Figure 4.3E, middle vs. right panel), despite similar test performances of both models (see Figures 4.3A&B). This observation is aligned with

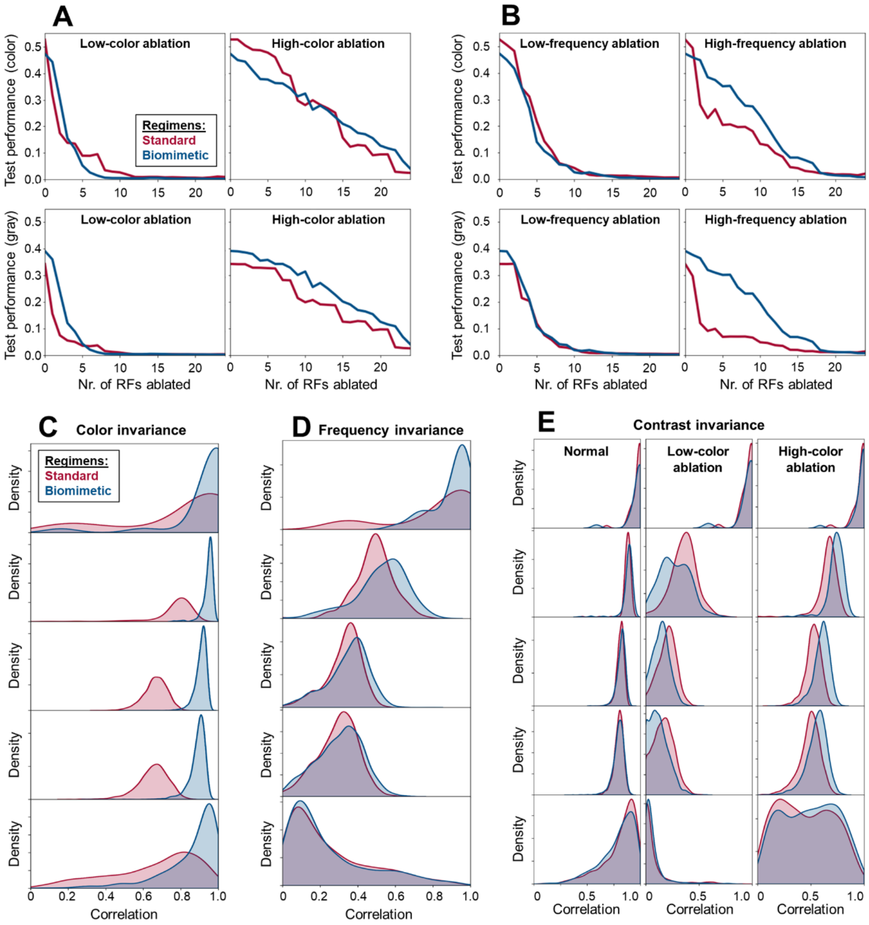


Figure 4.3: **A.** Classification performance on color (top) and grayscale (bottom) images when ablating the 24 (i.e., 50%) least (left) and most (right) colorful first-layer receptive fields. **B.** Classification performance on color (top) and grayscale (bottom) images when ablating the 24 (i.e., 50%) lowest (left) and highest (right) spatial frequency tuned receptive fields. **C&D.** Correlation of neural activations, across layers, between full-color and grayscale images (C) and full-frequency vs. blurred images (D). **E.** Correlation of neural activations, across layers, between full and reduced contrast images without ablation (left), when ablating the 50% least colorful receptive fields (middle), and when ablating the 50% most colorful receptive fields (right).

the neurophysiological finding that magno cells have greater contrast sensitivity and thereby saturate with lower contrast (Derrington and Lennie, 1984; Hicks et al., 1983; Kaplan and Shapley, 1982; Shapley et al., 1981), which would be expected to result in higher correlations between neural units across normal and low contrast stimuli.

4.5 DISCUSSION

We have presented an account of the genesis of the division of the parvo- and magnocellular pathways based on early developmental trajectories of sensory experience. Our computational results provide evidence in support of this account. Specifically, the results support the proposal that the joint coding of low spatial frequency and low color information in some receptive fields, and high spatial frequency and high color sensitivity in others, as characteristic of the division between the magnocellular and the parvocellular pathways, might be an outcome of the co-occurrence of these properties at different developmental time points. Further, in accordance with the expected psychophysical correlates of this division, we found that training with our biomimetic regimen not only induced more human-like classification decisions based on global shape rather than local texture information but also revealed the causal role of receptive fields exhibiting magnocellular characteristics in supporting such global shape bias. Similarly, the biomimetic network's classification performance is more differentially affected than that of the standard network by the ablation of low color or low spatial frequency units than by the ablation of high color or high spatial frequency units. The biomimetically-trained model also exhibited superior invariance in terms of its neural activations to the removal of chromatic or high spatial frequency content. Finally, the magnocellular-like receptive fields of the biomimetic model are potential drivers of contrast invariance, relative to the parvo-like receptive fields. On an applied note, the results also serve as a demonstration of how findings from biological development can help develop useful training protocols for computational systems (Zaadnoordijk et al., 2022).

It is important to note that while training with the biomimetic regimen resulted in a clear cluster of magnocellular receptive fields, units tuned to high spatial frequency or chromatic content exhibited greater heterogeneity. However, as previously pointed out, considering the mixing of the parvo and magno pathways at the level of the primary visual cortex (Hubel and Livingstone, 1990), resulting in two subpopulations of parvocellular units, the greater heterogeneity in the parvocellular relative to the magnocellular pathway is in agreement with neurophysiological reports. Further examining receptive field properties, an important question for future work is whether a biomimetic approach, as outlined in this paper, would be capable of also reproducing crucial characteristics of the parvo/magno distinction in the temporal domain, most notably the stronger sensitivity to rapid temporal changes of the magnocellular pathway (Derrington and Lennie, 1984; Dreher et al., 1976; O'Keefe et al., 1998; Schiller and Malpeli, 1978; Usrey and Reid, 2000).

Our computational results derived from the non-biomimetic control regimen, effectively dispensing with the initial phase of normal visual development, exhibit an important linkage to previous experimental findings in the domain of early visual deprivation. Specifically, in a reversible suture experiment on monkeys, Le Vay et al. (1980) found that artificially induced early deprivation, followed by later restoration of sight, yielded greater long-term damage to the magnocellular than the parvocellular units in the visual cortex. In light also of the earlier establishment (Rakic, 1977) and maturation (Gottlieb et al., 1985) of magno cells in the LGN, and the stronger input that magnocellular layers in the visual cortex receive at birth (Kennedy et al., 1985), the magno pathway's greater susceptibility to early deprivation has been accounted for by the earlier onset of its inputs. Human studies revealing late-sighted children's deficits in global motion processing (Ellemberg et al., 2002) also attest to the susceptibility of the magnocellular pathway to early visual deprivation, considering the task's alignment with magnocellular characteristics.

While the above magnocellular deficits might be accounted for by an anatomically hard-wired and pathway-specific critical period, following which, due to reduced plasticity, sight restoration does not allow the magnocellular pathway to gain normal function, an alternative account would be based not on the absence of sight during deprivation but on the quality of sight following sight restoration. This proposal is based on the observation that the normal maturational processes responsible for developmental improvements in perceptual aspects such as visual acuity continue to proceed despite the lack of visual inputs, such that the initial visual experience following sight-restoring surgery would be markedly richer (Boas et al., 1969; Hendrickson and Boothe, 1976). In other words, late-sighted individuals commence visual experience with high-quality inputs right from the start and thereby skip the phase of initially degraded vision characteristic of normal development. This, in turn, might prevent the magnocellular pathway from developing normally. Our computational simulations add plausibility to this idea, considering that training with high-quality visual inputs from the beginning, as opposed to initial training on low-frequency, achromatic stimuli, resulted in the marked absence of receptive fields exhibiting magnocellular characteristics. Future experimental work could add to this proposal.

Finally, the finding that training with a developmentally-inspired progression of inputs yields representations more consistent with empirical neurophysiological results, as reported in Figure 4.1, and yields more human-like global shape-based processing, as reported in Figure 4.2, is in keeping with results we previously reported in the domains of visual acuity (Vogelsang et al., 2018) and prenatal hearing (Vogelsang et al., 2023). In those studies, we found that initially low spatial acuity at birth and initially low temporal frequency sensitivity during prenatal development helped instantiate spatially or temporally extended receptive field structures and more robust performance profiles later in life. More generally, the principle that emerges from these studies is that initially degraded inputs appear to provide a scaffold rather than act as hurdles for the acquisition of later perceptual skills. In conclusion, the work presented here provides a potential account for the genesis of the parvo/magno distinction and thus also presents a teleological perspective on why normal development progresses in the way that it does. It further helps account for some of the impairments associated with atypical perceptual development and demonstrates how human development may help inspire training procedures of computational model systems.

ACKNOWLEDGEMENTS

This work has been supported by NIH grant R01EY020517 to PS.

DATA AND CODE AVAILABILITY

Data and code will be made available in a public repository upon publication.

4.6 SUPPLEMENTARY FIGURES

4.7 SUPPLEMENTARY METHODS

4.7.1 *Computational model*

For all simulations reported in this paper, we used the AlexNet CNN (Krizhevsky et al., 2012) and only made two minor adjustments to its architecture. First, the receptive fields in the first convolutional layer were enlarged (from a size of 11x11 pixels to 22x22 pixels) to avoid restricting the types of receptive field structures to be learned. Second, the number of receptive fields was reduced (from a total of 96 to 48) in order to reduce the proportion of overly noisy receptive fields, which would render the estimation of our metric distributions unreliable.

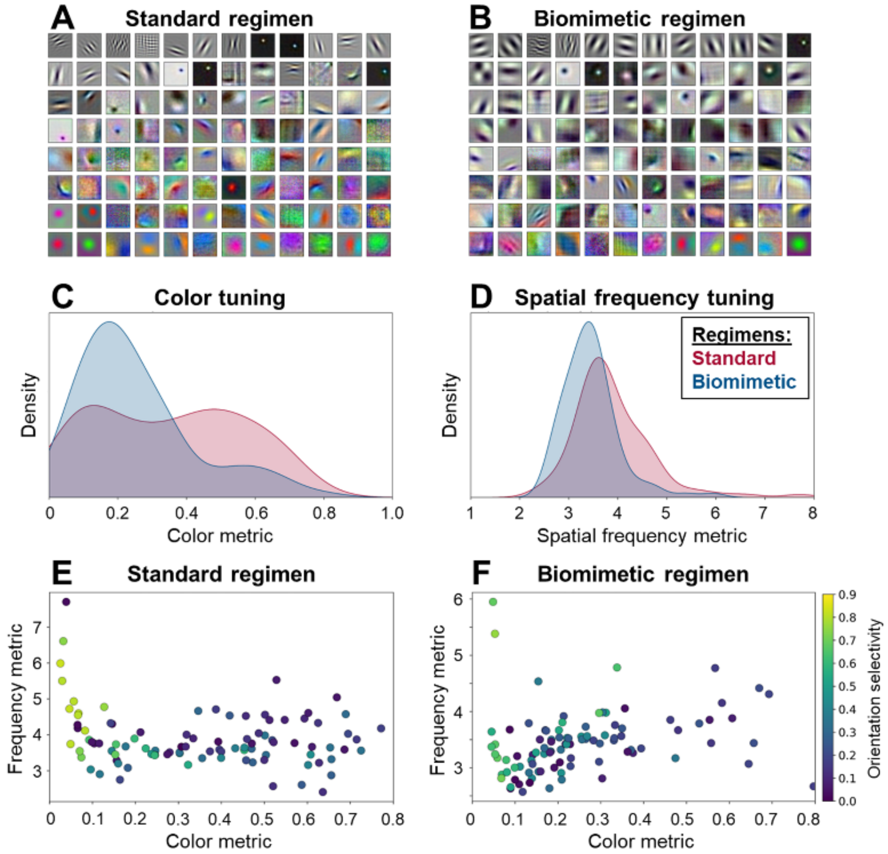


Figure 4.4: (Supplemental Figure) A reproduction of Figure 4.1 of the main manuscript when utilizing the classic AlexNet architecture comprised of 96, instead of 48, first-layer receptive fields. The results are qualitatively similar to those reported in the main manuscript.

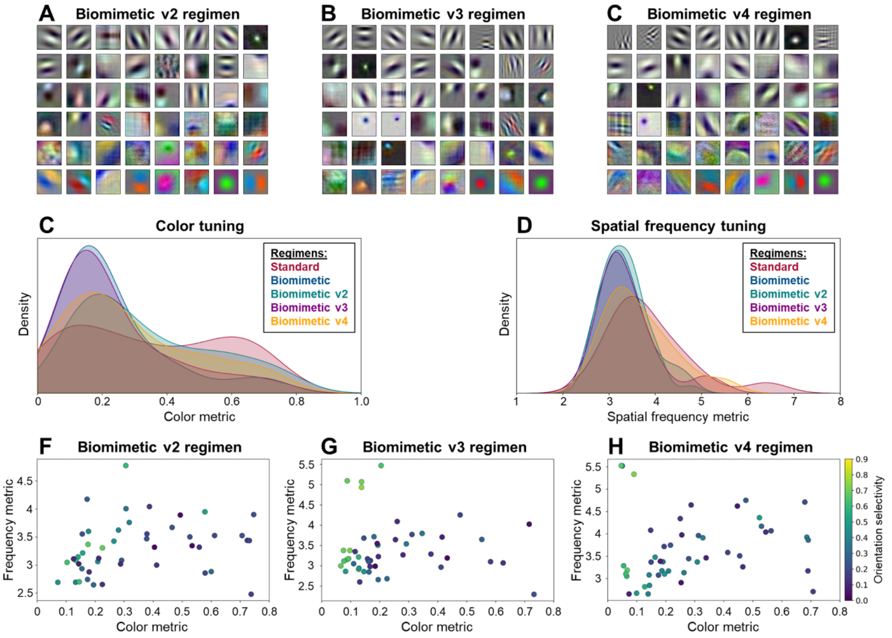


Figure 4.5: (Supplemental Figure) A reproduction of Figure 4.1 of the main manuscript when utilizing the additional biomimetic regimens v2-v4. The results are qualitatively similar to those reported in the main manuscript.

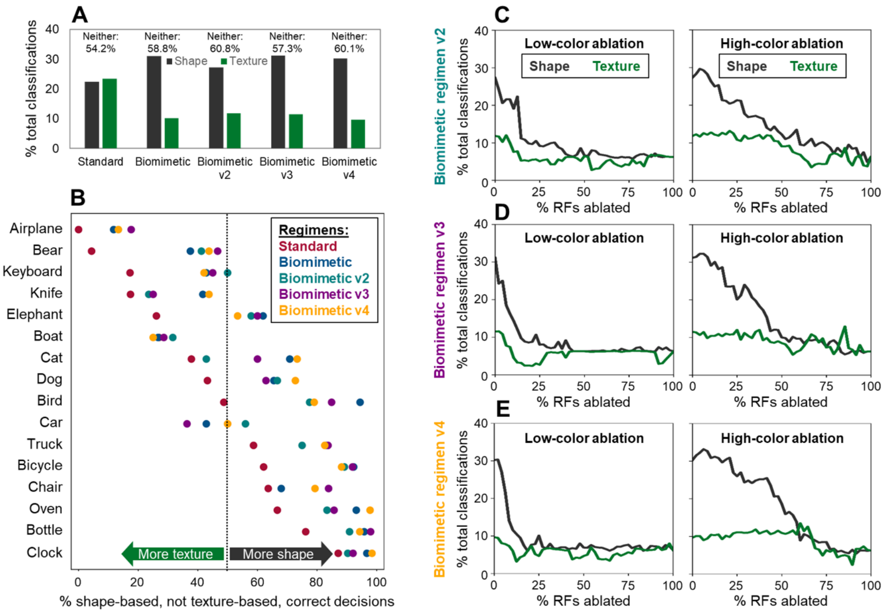


Figure 4.6: (Supplemental Figure) A reproduction of Figure 4.2 of the main manuscript when utilizing the additional biomimetic regimens v2-v4. The results are qualitatively similar to those reported in the main manuscript.

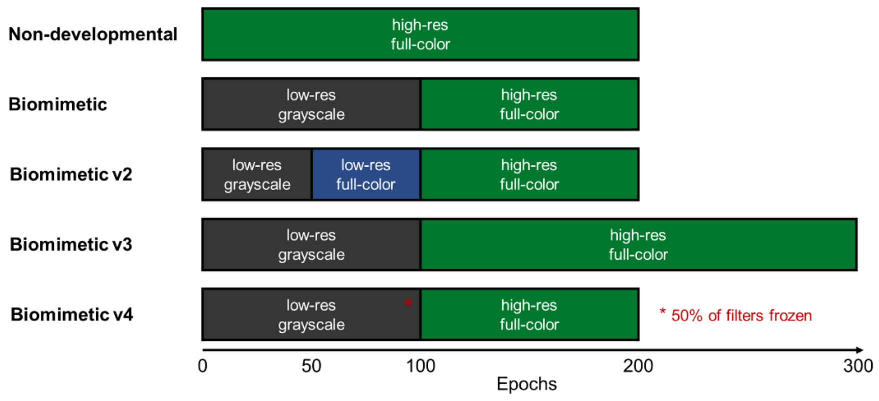


Figure 4.7: (Supplemental Figure) . Illustration of the total set of different training regimens used.

4.7.2 *Network training*

The slightly adjusted AlexNet was implemented and trained using Keras/TensorFlow v2. For training and testing, we utilized the official split of the ImageNet database (Deng et al., 2009) into a training set (containing more than 1 million images belonging to 1000 different object classes) and a test set (containing a total of 50,000 images – 50 for each object class). For training, we chose a batch size of 128, a constant learning rate of 0.001, categorical cross-entropy as loss function, and Stochastic Gradient Descent (SGD) as the optimizer, with a Nesterov momentum of 0.9. Image preprocessing and augmentation were kept fairly simple: random 227 x 227 segments were cropped out of the full 256 x 256 images, pixel values were rescaled from a $[0, 255]$ to a $[-1, 1]$ distribution, and images were flipped horizontally at random. Blurring (for the developmentally-inspired training regimens as well as for testing in Figure 4.3) was accomplished by applying a Gaussian blur with what would correspond to a sigma of 4.

4.7.3 *Different training regimens*

The above settings are generally applied to all training regimens used. The only variations were as follows (for illustration, see Supplemental Figure 4.7):

- In the ‘standard’ regimen, training lasted for a total of 200 epochs and contained exclusively high-resolution, full-color images.
- In the ‘biomimetic’ regimen, training on blurred, grayscale images for 100 epochs was followed by training on high-resolution, full-color images for 100 epochs.
- In the ‘biomimetic v2’ regimen, training on low-resolution, grayscale images for 50 epochs was followed by training on low-resolution,

full-color images for 50 epochs, and training on full-color, high-resolution images for 100 epochs.

- In the biomimetic v3 regimen, training on blurred, grayscale images for 100 epochs was followed by training on high-resolution, full-color images for 200 epochs.
- In the biomimetic v4 regimen, training on blurred, grayscale images for 100 epochs was followed by training on high-resolution, full-color images for 100 epochs (as in the 'biomimetic' regimen) but a proportion (here, 50%) of the first-layer filters was confined to only learn during the second half of training.

4.7.4 *Color metric*

We quantified the colorfulness of a given receptive field in two steps. First, we extracted the intensity differences, m, across the R, G, and B color channels for each individual pixel:

$$\begin{aligned} x &= R\cos(0) + G\cos\left(\frac{2}{3}\pi\right) + B\cos\left(-\frac{2}{3}\pi\right) \\ y &= R\sin(0) + G\sin\left(\frac{2}{3}\pi\right) + B\sin\left(-\frac{2}{3}\pi\right) \\ m &= \sqrt{x^2 + y^2} \end{aligned}$$

R, G, and B thereby represent the pixel-by-pixel channel intensities. We then summarized the distribution of such color channel imbalances across the 22x22 pixels of a given receptive field into a single value. Defining our final color metric as the mean of the 22x22 distribution would induce an estimation bias towards spatially extended color receptive fields. Instead, taking into account only the single most colorful pixel would be subject to high noise. As a compromise, we chose to define our final metric as the average of the top-48 (i.e., approximately the top-10%) most colorful pixels within a receptive field – roughly approximating the size of the smallest effective receptive fields.

4.7.5 *Spatial frequency metric*

To measure spatial frequency content, we applied a 2D-FFT to a grayscaled version of each receptive field and used radial averaging to summarize the presence of different spatial frequencies, providing us with a 1D histogram over frequencies. Given such histogram, we defined our spatial frequency metric as the weighted average frequency:

$$\textbf{weighted average frequency} = \frac{\sum_f amp(f) * f}{\sum_f amp(f)}$$

Amp thereby refers to the amplitude of a given frequency, and f refers to the frequency itself. There are a few details worth pointing out. First, note that the constant part of the FFT was excluded for the calculation. Further, in order to avoid any noise that may be caused by the discreteness of the index, all receptive fields (natively 22x22 pixels) have been up-sampled by a factor of 100 prior to applying the 2D-FFT. Finally, note that the metric is independent of the absolute strength of the signal, rendering the question of whether to use a normalized or unnormalized spectral decomposition obsolete.

4.7.6 *Orientation selectivity metric*

Finally, we wished to define a metric to capture the extent of orientation tuning exhibited by individual receptive fields. Akin to the computation of the spatial frequency metric, we started by carrying out a 2D-FFT on each receptive field. As opposed to radial averaging, where direction-independent frequency profiles are extracted, we here applied azimuthal averaging, where frequency bands are averaged across. Note that the utilized frequencies were restricted to only a quarter of the theoretically available frequency spectrum – a compromise chosen in order to filter out some high-frequency noise while not removing any effective frequencies in the receptive fields. Overall, this allows us to assess intensity as a function of orientation, ranging from zero to π. Orientation-tuned receptive fields are associated with a sharp and strong peak in this distribution. This can be captured well by the mean resultant length (Fisher, 1995):

$$R = \frac{1}{\sum_\theta f(\theta)} \left| \sum_\theta f(\theta) e^{2i\theta} \right|$$

where θ represents the orientations from 0 to π, and $f(\theta)$ their corresponding azimuthally-averaged amplitudes. A line plot depicting the intensity over orientation from 0 to π can thereby be imagined to be plotted around a whole circle, with the origin of the 2D coordinate system in the center and orientations 90 degrees apart from each other on opposing ends of the circle. Our metric then simply represents the length of the vector resulting from averaging in this coordinate system.

4.7.7 *Texture/shape analysis*

For the results reported in Figure 4.2 (and Supplemental Figure 4.6), we have used the test images from Geirhos et al. (2018), which are provided online. These are 1280 images, each with a shape-texture conflict. For instance, an image may have the global shape of an airplane but the local texture of a cat. If the network classifies such image as an airplane, the decision would be judged as shape-consistent; if it classifies the image as a cat, it would be judged as texture-consistent; and otherwise, it would be judged as incorrect (depicted as 'neither' in Figure 4.2 and Supplemental Figure 4.6).

4.7.8 *Invariance analysis*

For the invariance analysis reported in Figures 4.3C-E, we computed the correlations between unit activations, which were concatenated for 3000 test set items (3 images per ImageNet class) for three separate image manipulations we carried out: full-color vs. grayscale (Figure 4.3C), full-frequency vs. blur (using a Gaussian blur with a sigma of 4) (Figure 4.3D), and full-contrast vs. reduced contrast (with a contrast reduction factor of 0.5) (Figure 4.3E). The correlation was thereby calculated for each unit flattened over the dimensions and all the test images.

REFERENCES

Adams, Russell J and Mary L Courage (2002). "A psychophysical test of the early maturation of infants' mid-and long-wavelength retinal cones." In: *Infant Behavior and Development* 25.2, pp. 247–254.

Banks, Martin S and Patrick J Bennett (1988). "Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates." In: *JOSA A* 5.12, pp. 2059–2079.

Boas, Judith AR, RL Ramsey, AH Riesen, and JP Walker (1969). "Absence of change in some measures of cortical morphology in dark-reared adult rats." In: *Psychonomic Science* 15.5, pp. 251–252.

Candy, T Rowan and Martin S Banks (1999). "Use of an early nonlinearity to measure optical and receptor resolution in the human neonate." In: *Vision research* 39.20, pp. 3386–3398.

Courage, Mary L and Russell J Adams (1990). "Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples." In: *Optometry and vision science* 67.9, pp. 713–718.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei (2009). "Imagenet: A large-scale hierarchical image database." In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Derrington, AM and P Lennie (1984). "Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque." In: *The Journal of physiology* 357.1, pp. 219–240.

Dobkins, Karen R, Barry Lia, and Davida Y Teller (1997). "Infant color vision: Temporal contrast sensitivity functions for chromatic (red/-green) stimuli in 3-month-olds." In: *Vision Research* 37.19, pp. 2699–2716.

Dobson, Velma and Davida Y Teller (1978). "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies." In: *Vision research* 18.11, pp. 1469–1483.

Dreher, B, Y Fukada, and RW Rodieck (1976). "Identification, classification and anatomical segregation of cells with X-like and Y-like properties in the lateral geniculate nucleus of old-world primates." In: *The Journal of Physiology* 258.2, pp. 433–452.

Ellemberg, Dave, Terri L Lewis, Daphne Maurer, Sonia Brar, and Henry P Brent (2002). "Better perception of global motion after monocular than after binocular deprivation." In: *Vision research* 42.2, pp. 169–179.

Fisher, Nicholas I (1995). *Statistical analysis of circular data*. cambridge university press.

Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel (2018). "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness." In: *arXiv preprint arXiv:1811.12231*.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press.

Gottlieb, Michael D, Pedro Pasik, and Tauba Pasik (1985). "Early postnatal development of the monkey visual system. I. Growth of the lateral geniculate nucleus and striate cortex." In: *Developmental Brain Research* 17.1-2, pp. 53–62.

Hendrickson, Anita and Ronald Boothe (1976). "Morphology of the retina and dorsal lateral geniculate nucleus in dark-reared monkeys (Macaca nemestrina)." In: *Vision research* 16.5, 517–IN5.

Hicks, TP, BB Lee, and TR Vidyasagar (1983). "The responses of cells in macaque lateral geniculate nucleus to sinusoidal gratings." In: *The Journal of physiology* 337.1, pp. 183–200.

Hubel, David H and Margaret S Livingstone (1990). "Color and contrast sensitivity in the lateral geniculate body and primary visual cortex of the macaque monkey." In: *Journal of neuroscience* 10.7, pp. 2223–2237.

Jacobs, Deborah S and Colin Blakemore (1988). "Factors limiting the postnatal development of visual acuity in the monkey." In: *Vision research* 28.8, pp. 947–958.

Kaplan, E and RM Shapley (1982). "X and Y cells in the lateral geniculate nucleus of macaque monkeys." In: *The Journal of Physiology* 330.1, pp. 125–143.

Kennedy, H, J Bullier, and C Dehay (1985). "Cytochrome oxidase activity in the striate cortex and lateral geniculate nucleus of the newborn and adult macaque monkey." In: *Experimental Brain Research* 61.1, pp. 204–209.

Kiorpes, Lynne and J Anthony Movshon (2004). "Neural limitations on visual development in primates." In: *The visual neurosciences* 1, pp. 159–173.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems* 25.

Le Vay, Simon, Torsten N Wiesel, and David H Hubel (1980). "The development of ocular dominance columns in normal and visually deprived monkeys." In: *Journal of Comparative Neurology* 191.1, pp. 1–51.

Lindsay, Grace W (2021). "Convolutional neural networks as a model of the visual system: Past, present, and future." In: *Journal of cognitive neuroscience* 33.10, pp. 2017–2031.

Livingstone, Margaret S and David H Hubel (1987). "Psychophysical evidence for separate channels for the perception of form, color, movement, and depth." In: *Journal of Neuroscience* 7.11, pp. 3416–3468.

Livingstone, Margaret S, Glenn D Rosen, Frank W Drislane, and Albert M Galaburda (1991). "Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia." In: *Proceedings of the National Academy of Sciences* 88.18, pp. 7943–7947.

Livingstone, Margaret and David Hubel (1988). "Segregation of form, color, movement, and depth: anatomy, physiology, and perception." In: *Science* 240.4853, pp. 740–749.

O'Keefe, Lawrence P, Jonathan B Levitt, Daniel C Kiper, Robert M Shapley, and J Anthony Movshon (1998). "Functional organization of owl monkey lateral geniculate nucleus and visual cortex." In: *Journal of neurophysiology* 80.2, pp. 594–609.

Rakic, Pasko (1977). "Genesis of the dorsal lateral geniculate nucleus in the rhesus monkey: site and time of origin, kinetics of proliferation, routes of migration and pattern of distribution of neurons." In: *Journal of Comparative Neurology* 176.1, pp. 23–52.

Schiller, Peter H and Joseph G Malpeli (1978). "Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey." In: *Journal of neurophysiology* 41.3, pp. 788–797.

Schrimpf, Martin, Jonas Kubilius, Michael J Lee, N Apurva Ratan Murty, Robert Ajemian, and James J DiCarlo (2020). "Integrative benchmarking to advance neurally mechanistic models of human intelligence." In: *Neuron* 108.3, pp. 413–423.

Shapley, Robert (1990). "Visual sensitivity and parallel retinocortical channels." In: *Annual review of psychology* 41.1, pp. 635–658.

Shapley, Robert (1992). "Parallel retinocortical channels: X and Y and P and M." In: *Advances in psychology*. Vol. 86. Elsevier, pp. 3–36.

Shapley, Robert, Ehud Kaplan, and Robert Soodak (1981). "Spatial summation and contrast sensitivity of X and Y cells in the lateral geniculate nucleus of the macaque." In: *Nature* 292.5823, pp. 543–545.

Usrey, W Martin and R Clay Reid (2000). "Visual physiology of the lateral geniculate nucleus in two species of New World monkey:

Saimiri sciureus and Aotus trivirgatis." In: *The Journal of Physiology* 523.3, pp. 755–769.

Vogelsang, Lukas, Sharon Gilad-Gutnick, Evan Ehrenberg, Albert Yonas, Sidney Diamond, Richard Held, and Pawan Sinha (2018). "Potential downside of high initial visual acuity." In: *Proceedings of the National Academy of Sciences* 115.44, pp. 11333–11338.

Vogelsang, Marin, Lukas Vogelsang, Sidney Diamond, and Pawan Sinha (2023). "Prenatal auditory experience and its sequelae." In: *Developmental Science* 26.1, e13278.

Wiesel, Torsten N and David H Hubel (1966). "Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey." In: *Journal of neurophysiology* 29.6, pp. 1115–1156.

Zaadnoordijk, Lorijn, Tarek R Besold, and Rhodri Cusack (2022). "Lessons from infant learning for unsupervised machine learning." In: *Nature Machine Intelligence* 4.6, pp. 510–520.

5 BUTTERFLY EFFECTS IN PERCEPTUAL DEVELOPMENT: A REVIEW OF THE 'ADAPTIVE INITIAL DEGRADATION' HYPOTHESIS

Content from

Vogelsang, L.*, **Vogelsang, M.***, Pipa, G., Diamond, S., & Sinha, P. (Submitted). "Butterfly effects in perceptual development: a review of the 'adaptive initial degradation' hypothesis". **Under peer-review** at Developmental Review (initial submission: May 2023). (* = equal contribution)

5.1 ABSTRACT

Human perceptual development evolves in a stereotyped fashion, with initially limited perceptual capabilities maturing over the months or years following commencement of sensory experience into robust proficiencies. This review focuses on the functional significance of these developmental progressions. Specifically, we review findings from studies of children who have experienced alterations of early development, as well as results from corresponding computational models, which have recently provided compelling evidence that specific attributes of early sensory experience are likely to be important prerequisites for later developing skills in several perceptual domains such as vision and audition. Notably, the limitations of early sensory experience have therein emerged as scaffolds, rather than hurdles, being causally responsible for the acquisition of later perceptual proficiencies, while dispensing with these limitations has the perhaps counter-intuitive consequence of compromising later development. These results have implications for understanding why normal trajectories of perceptual development are sequenced in the way that they are, help account for the perceptual deficits observed in individuals with atypical histories of sensory development, and serve as guidelines for the creation of more robust and effective training procedures for computational learning systems.

5.2 KEYWORDS

adaptive initial degradations; visual development; late sight-onset; prenatal hearing; deep networks

5.3 INTRODUCTION

Many aspects of human perceptual development exhibit a consistent choreography in terms of how they unfold over time. A typically-developing infant starts out with limited perceptual capabilities at birth and progressively acquires greater proficiencies over the ensuing months and years. For instance, in the domain of visual perception, the development of contrast sensitivity, resolution acuity, color sensitivity, and the presence of neural noise (Dobkins et al., 1997; Kiorpes, 2016; Lenassi et al., 2008; Skoczenski and Norcia, 1998) illustrates how initially limited perceptual capabilities mature over the months and years following birth into robust proficiencies.

5.3.1 *Two views on the functional significance of early developmental limitations*

Our focus in this paper is to reflect on the functional significance, if any, of these developmental progressions, evolving from limited to proficient. There are two contrasting perspectives to consider. The first and more traditional one treats such progressions purely as epiphenomena that accompany maturation. In this view, early perceptual limitations are inevitable outcomes of the physiological immaturities of the underlying sensory/neural cells and circuits. Amelioration of these immaturities over time leads, unsurprisingly, to an improvement in perceptual proficiencies. In keeping with this view, the acquisition of functional skills needs to surmount the challenges imposed by the early limitations in order to attain later manifesting proficiencies.

In contrast to this perspective is the more recent proposal that initial perceptual limitations may not be hurdles, but rather act as scaffolds. Instead of merely being secondary consequences of physiological maturation, they may play a primary role in causally shaping subsequent development, by providing a drive to instantiate processing mechanisms that prove to be beneficial later in life. In other words, the temporal sequencing of developmental stages, from limited to proficient, may, in fact, serve an adaptive purpose; later perceptual proficiencies may arise not despite the early limitations, but in part because of them. To loosely borrow Lorenz's (Lorenz, 1963) metaphor from which this article derives its title, early perceptual limitations may be akin to the flapping of a butterfly's wings, setting up small eddies that manifest in due time as significant salutary effects on later perceptual skills. We call this the 'Adaptive Initial Degradation' (AID) hypothesis.

5.3.2 *Past support for initial degradations during development being adaptive*

While the first viewpoint has historically been the dominant one in the field, some past evidence has lent credence to the second perspective. Part of this evidence is not directly from the domain of development but nevertheless serves to motivate this perspective. This includes work showing the effectiveness of coarse processing to facilitate fine analysis. For instance, a coarse-to-fine approach to disparity detection was found to be a useful strategy for the task of stereo-correspondence (Sizintsev and Wildes, 2010). Similarly, in the context of optic flow estimation, coarse-to-fine methods have been found to yield higher accuracy (Anandan, 1989; Black and Anandan, 1996; Mémin and Pérez, 2002). Exhibiting a more direct link to development, Turkewitz and Kenny (1982) were among the first to propose that commencing sensory experience with initially simpler stimuli renders perceptual analysis less overwhelming, thereby supporting, rather than hindering, the acquisition of perceptual proficiencies. Further, Elman (1993) and Newport (1988) proposed, and drew on computational simulations to support their proposals, that language learning benefits from early developmental limitations in cognitive architectures. Similarly, Dominguez and Jacobs (2003) demonstrated that the acquisition of binocular disparity detection is supported by the immaturity of an infant's visual system. These earlier studies help motivate the need for a more comprehensive examination of the role of developmental progressions in the context of the AID hypothesis.

5.3.3 *Two recently emerged research avenues for examining the AID hypothesis*

Recently, two promising avenues for more directly probing the potential functional significance of initial developmental limitations have emerged. The first lies in the assessment of perceptual skills of children who have experienced alterations of early developmental experience relative to their typically-developed peers. These alterations include, for instance, periods of visual deprivation due to congenital blindness and truncations of in-utero auditory experience due to premature birth. These cases, although rare, present an unusual opportunity to examine how an altered developmental progression impacts later perceptual skills. For instance, individuals who are born blind and gain sight later in life commence their visual experience with sensory and neural mechanisms that are more mature than a neonate's. Hence, their developmental trajectories effectively dispense with the initial phase of degradations characteristic of normal development. Examining the perceptual profiles of such individuals can be useful for assessing the validity of the AID hypothesis, i.e., whether early degradations part of typical development may serve an adaptive purpose and whether their dispensing with would result in subsequent perceptual deficits.

A second avenue that has emerged even more recently, but has already proven its utility in the developmental domain, lies in the systematic experimentation with a variety of modern computational model systems – most commonly, deep neural networks. While these networks are not exact models of their biological counterparts, they are among the most successful models for predicting human behavior and neural responses across the sensory hierarchy (Cadena et al., 2019; Lindsay, 2021; Schrimpf et al., 2020; Storrs et al., 2021) and can thereby serve as useful approximations of sensory processing mechanisms. Crucially, these systems offer a systematic methodology for directly examining the consequences of controlled manipulations of sensory experience that experiments with humans, for both ethical and practical reasons, do not provide. Specifically, one can expose (or 'train') a deep neural network, whose processing machinery is formed (or 'learned') directly in response to the inputs it is exposed to, on several temporal progressions of sensory inputs. While some of these can be chosen to be 'biomimetic' in the sense that it recapitulates certain aspects of typical development (e.g., comprising the presentation of visual inputs that transition from initial blurring to later high-resolution), others can serve as non-developmental controls (e.g., sets of visual stimuli incorporating exclusively high-resolution images from the start of network training). Following exposure to these different types of experiential history, one can then examine the consequences on the systems' performance on defined perceptual tasks, akin to how perceptual proficiencies of typically- and atypically-developed children are being examined. In addition, one can also probe the consequences of the different training regimens on the learned inner workings (or 'representations') of the networks, which can be quantitatively related to cortical processing in the biological system (Kriegeskorte, 2015; Schrimpf et al., 2020).

Taken together, these empirical and computational studies can fruitfully complement each other and help to broadly establish the potential functional significance of initial limitations in sensory development for the acquisition of later perceptual skills.

5.3.4 *Aim and structure of this paper*

Recent findings from both of the aforementioned empirical and computational studies, from our lab and others, have provided compelling evidence in favor of the AID hypothesis, attesting to specific attributes of early sensory experience likely being important prerequisites for later developing skills in several perceptual domains. Notably, the 'limitations' of early sensory experience have therein emerged as being causally responsible for later proficiencies, whereas dispensing with these limitations has the perhaps counter-intuitive consequence of compromising later development. Here, we focus on reviewing and discussing these findings to examine the plausibility of the adaptive initial degradation hypothesis along several visual and auditory perceptual dimensions. For each dimension, we briefly describe the basic developmental progression, a hypothesis for how the initial limitations might be adaptive, and, whenever applicable, past behavioral and/or computational tests of these.

5.4 AN EXAMINATION OF THE AID HYPOTHESIS IN THE CONTEXT OF VISUAL ACUITY

5.4.1 *The developmental progression of visual acuity in typically-developing and late-sighted individuals*

Typically-developing infants begin to experience the visual world with strikingly poor acuity – below 20/600 (Courage and Adams, 1990; Dobson and Teller, 1978) – which is mostly attributable to immaturities in the retina (Banks and Bennett, 1988; Yuodelis and Hendrickson, 1986) and visual cortex (Jacobs and Blakemore, 1988; Kiorpes and Movshon, 2004). Throughout the developmental time course, these immaturities steadily diminish, resulting in 20/20 vision after a few years of life (Daw, 2014).

This developmental trajectory, characterized by steady improvements in visual acuity, markedly differs from that of a child that is born with bilateral cataracts and gaining sight later in life. For such individual, the retina and visual cortex keep maturing despite the lack of visual inputs, such that upon eventual cataract removal, her initial acuity is far exceeding that of a newborn (Banks and Crowell, 1993; Boas et al., 1969; Hendrickson and Boothe, 1976; Wilson, 1993).

5.4.2 *A hypothesis for how initially low visual acuity may be adaptive*

Although it might intuitively appear suboptimal to commence visual experience with poor acuity, given the weaker stimulus quality it offers, the following provides a rationale, in accordance with the AID hypothesis, for how it might, in fact, be beneficial. Significantly reduced acuity, as is characteristic of normal development, effectively induces visual blur. Small segments in blurry images do not carry enough sensory information to allow for informative inference upon the external environment. To nevertheless be able to detect meaningful patterns in the visual world, integration over spatially extended areas is necessary (Kwon et al., 2016). To the contrary, having access to high-resolution imagery right from the beginning would render the instantiation of such spatial integration mechanisms superfluous, considering that even local processing would suffice for visual dis-

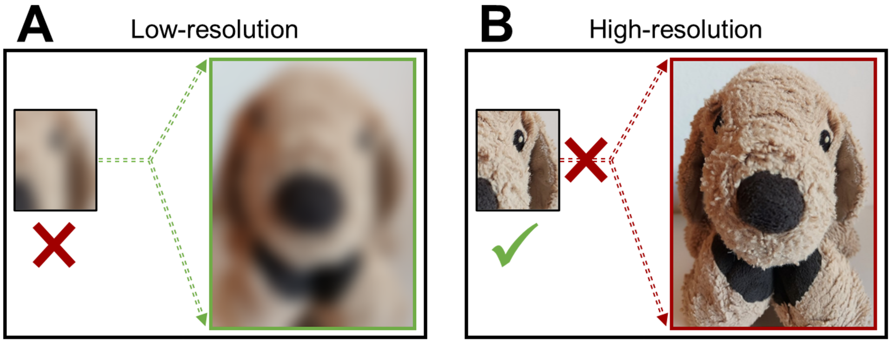


Figure 5.1: An intuition for the potential benefits of commencing visual experience with initially reduced acuity. **A.** A small segment of a blurry image proves insufficient for visual discrimination. It therefore necessitates integration across larger spatial extents. **B.** A small segment of a high-resolution image may well suffice for visual classification, thereby rendering the development of extended spatial integration superfluous

crimination in such circumstances (Ince et al., 2015; Smith and Schyns, 2009). This intuition is illustrated in Figure 5.1.

Stated differently, the poor initial visual acuity characteristic of normal development might provide a basis for inducing extended spatial integration mechanisms that, consistent with the AID hypothesis, would facilitate perceptual analyses later in life. Following the AID hypothesis, commencing visual experience with abnormally high acuity instead, as it happens with late-sighted individuals, might result in compromised spatial integration and, consequently, in reduced performance on tasks crucially reliant on it.

5.4.3 *Evidence from empirical studies with late-sighted individuals*

Several years ago, we have had the unusual opportunity to test the above hypothesis through our meeting with a young boy, RK, from China. RK had been born blind, with dense cataracts, and was placed in an orphanage soon thereafter, not able to receive treatment until the age of 4.5 years. Two years after his surgery, an American family adopted him and observed that despite being otherwise visually proficient, RK kept having difficulty recognizing people's faces. Upon the parents' request, we invited RK to our lab at MIT and, in accordance with their observations, found that he performed well on a variety of basic visual tests except for face identification (Vogelsang et al., 2018). A review of the literature revealed that RK's specific face recognition deficit is consistent with results that have previously been reported in other children who had undergone treatments for congenital cataracts (De Heering and Maurer, 2014; Geldart et al., 2002; Putzar et al., 2010), notably even when treatment took place already within the first year of life.

These findings, attesting to early visual deprivation leading to deficits in facial identification while other basic visual proficiencies appear to not be significantly affected, can well be accounted for by the AID hypothesis. Specifically, extended spatial integration is known to be a requirement for the detection of configural visual relationships, which play an especially important role in the identification of faces (Goffaux et al., 2005; Peterson and Rhodes, 2003; Richler and Gauthier, 2014; Taubert et al., 2011; Young et al., 1987). Considering that, due to delayed onset of sight, the initial visual acuity is abnormally high (in the case of RK, he began seeing with 20/40, instead of 20/600, vision), the AID hypothesis predicts that the development of extended spatial integration is compromised, resulting in reduced performance on such tasks as face identification that rely on it heavily, while not affecting more basic visual tasks. These predictions were empirically confirmed (Vogelsang et al., 2018).

It is important to note that the selective impairment of facial identification has previously also been accounted for as the manifestation of a face-specific critical period (Rivolta, 2014; Röder et al., 2013) – the idea that the perceptual and cortical specialization needed for the emergence of proper face identification abilities requires exposure to faces early in life and cannot be made up for later (Arcaro et al., 2017; De Schonen and Mathivet, 1989; Geldart et al., 2002; Nelson, 2001). While such an account cannot be entirely ruled out on the basis of the above experiments alone, the AID hypothesis does not assume the effect to be specific to the domain of faces, and therefore has the advantage of parsimony. Future experimentation on the potential neural mechanisms underlying such behavior (see next section) as well as further empirical studies investigating the performance of late-sighted individuals on such additional tasks requiring spatial integration as contour integration (Field et al., 1993), motion coherence (Newsome and Pare, 1988), and symmetry detection (Dakin and Herbert, 1998), can contribute to further disentangling these two theoretical proposals.

5.4.4 *Evidence from computational simulations*

As noted in the introductory section of this paper, systematic experimentation on deep neural networks as computational model systems can add support of the AID hypothesis to that derived from empirical studies of late-sighted individuals. Key to their utility is that they allow for the deliberate manipulation of experiential history and a systematic examination of the system's resulting inner workings and perceptual proficiency. To this end, we had conducted a computational study, as detailed in Vogelsang et al. (2018), in which we exposed different instances of a deep convolutional neural network (Krizhevsky et al., 2012) to a dataset of face images (Ng and Winkler, 2014), while systematically varying the amount of blur applied to the images throughout training. Specifically, motivated by, and serving as a rough proxy of, biological development, we utilized a 'biomimetic' regimen, in which we exposed the network to blurred images for the first half of training, followed by exposure to high-resolution imagery for the second half (we term this the 'low-to-high' regimen). To allow for a comprehensive comparison, we also trained three additional instances of the network using 'non-biomimetic' control regimens: 'high-to-low' (training on high-resolution followed by low-resolution images, i.e., following an inverse-developmental order), 'low-only' (training on exclusively low-resolution images), and 'high-only' (training on exclusively high-resolution ones). We then examined two aspects of the differently trained network instances. First, we evaluated their 'receptive fields' (RFs) in the first convolutional layer, which are in rough analogy to receptive fields found in the visual cortex and whose shapes, which are sculpted in response to the stimuli that the network is exposed to during training, are indicative of the spatial extent with which visual features are integrated. Second, we tested the four trained network instances' ability to identify faces, considering the reliance of this ability on spatially extended analyses (Peterson and Rhodes, 2003; Young et al., 1987), across different levels of blur.

Examining the networks' learned receptive fields, we found that exposure to exclusively low-resolution images, relative to exposure to exclusively high-resolution images, resulted in markedly extended receptive field structures (see Figures 5.2A), indicative of the network's use of more long-range spatial relationships. Interestingly, as is evident in the biomimetic 'low-to-high' regimen, the large RFs established through initial training with blurred imagery did not exhibit any shrinkage when later exposed to high-resolution ones. This is in stark contrast to the inverse-biomimetic training regimen, where RF sizes increased strongly when high-resolution training was followed by training with blurred imagery (see Figure 5.2B). This marked effect of ordering indicates that initially large RFs may represent more stable and more generally useful processing blocks, therefore obviating the need for their later adjustment. More closely examining training protocols in which initial training on blurry images was followed by training on high-resolution imagery, we further found that even short periods of exposure to initially blurry imagery (here, around 20% of the entire training phase) were sufficient to induce stably large RFs (see Figure 5.2C). Thus, even the short period of low acuity that humans exhibit early in infancy might well suffice for the establishment of stably large receptive fields.

Further evaluation of network performance on a face identification task, tested on high-resolution images as well as others with several different degrees of blurring, revealed remarkable differences between training regimens. Most notably, the biomimetic model (i.e., the one trained with the developmentally-inspired 'low-to-high' regimen) pro-

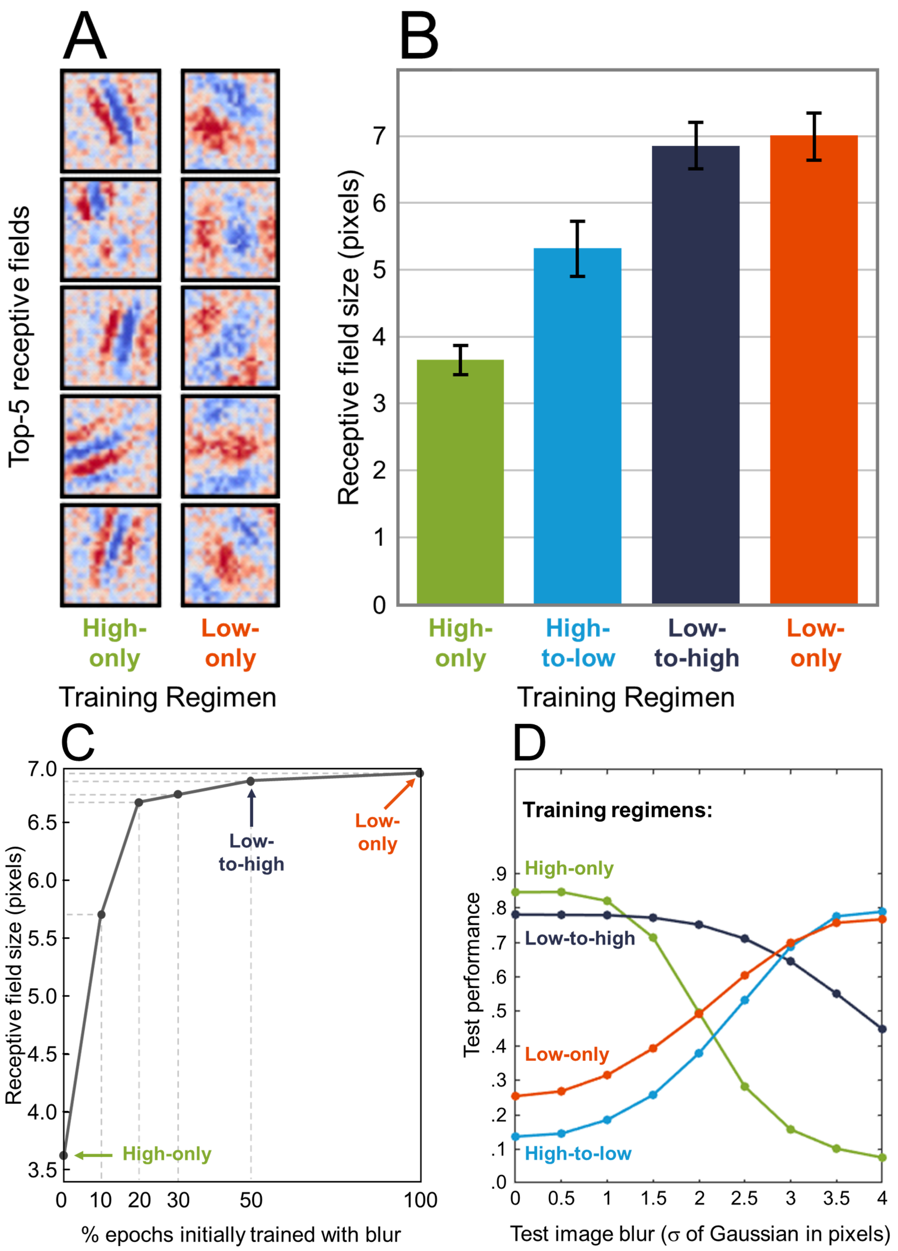


Figure 5.2: **A.** Receptive fields emerging from training on high-only and low-only regimens. Low-acuity experience leads to larger RFs. **B.** Impact of different training paradigms on RF sizes. The low-to-high model exhibits almost no shrinkage in RF size, relative to the low-only model **C.** When training begins with blur and proceeds with high-resolution imagery, having as little as 20% of the whole training be on blur proves sufficient for the system to produce stably large RFs **D.** Impact of different training paradigms on performance: the low-to-high model exhibits the most generalized performance. These plots were adapted from Vogelsang et al. (2018).

duced the most superior and generalized classification performance across all resolution levels, by correspondence to the largest area under the curve and lowest decay as a function of blur (see Figure 5.2D, black curve). To the contrary, the inverse-developmental regimen ('high-to-low'; blue curve) resulted in especially poor generalization across blur, even though it had been trained with the same set of images as the biomimetic one, only in reverse order. Thus, the superior performance of the low-to-high models appears to be a direct consequence of the initial exposure to degraded imagery, akin to the sensory experience of typically-developed infants.

These computational results, published in Vogelsang et al. (2018), lend additional support to the AID account and the proposal that initial experience with degraded imagery may help set up receptive fields capable of encoding image structure over extended spatial extents. This, in turn, helps reduce reliance on local features and improves generalization performance across a range of image resolutions.

5.4.5 *Further computational studies*

The publication of the work presented above has elicited several responses. For instance, Jinsi et al. (2023), in support of the HIA proposal, demonstrated that gradual low-to-high-resolution training regimens improve basic-level as well as subordinate-level categorization performance, therefore attesting to the functional significance conferred by training with initially degraded visual inputs. In addition, Jang and Tong (2021) confirmed the findings of blurry-to-high-resolution training inducing greater robustness to blur for the task of face recognition. However, they did not find this specific effect in the context of broader object recognition, thereby highlighting potential differences in the holisticness of processing between the two types of tasks. Further, although not in the specific context of developmental progressions, Kong et al. (2022) showed that convolutional neural networks that developed comparatively lower spatial frequency preferences were more aligned with the spatial frequency tuning of cells found in the primary visual cortex of macaque monkeys, and also induced the model to exhibit greater robustness to image perturbations; leaving it to future work to examine the specific relationship to developmental trajectories commencing with initially low acuity. Finally, it is worth pointing out that, as highlighted in the response to the article of Vogelsang et al. (2018) by Katzhendler and Weinshall (2019), the generalization performance of the low-to-high-resolution regimen, although markedly superior to those of the high-to-low, low-only, or high-only regimens, does not fully reach the level of a brute-force data augmentation training regimen in which blur levels would be mixed throughout training. However, as commented on in Vogelsang et al. (2019), such mixed regimens, though straightforward to probe computationally, do not appear to be easily implementable biologically. As such, the low-to-high-resolution trajectory observed in typical development might be the most suitable compromise between biological feasibility and generalization performance. Thus, the inferiority of developmentally-inspired training to fully augmented mixed regimens does not invalidate the AID account but rather attests to the remaining pieces of the puzzle of bringing developmental insights into machine learning practice.

5.5 EXAMINING THE AID HYPOTHESIS BEYOND THE VISUAL DOMAIN: PRENATAL EXPERIENCE IN THE AUDITORY MODALITY

5.5.1 *The developmental progression of temporal frequency sensitivity in prenatal hearing*

Beyond the domain of vision, the AID hypothesis may apply across additional key dimensions of sensory development. A particularly striking case is the auditory analog of visual acuity, in which the developmental progression from limited to proficient does not transpire over the first years of life but, in fact, already begins prenatally. Specifically, by around 20 weeks of gestational age, a human fetus begins to be able to perceive voices and other sounds in the mother's external environment (Gerhardt and Abrams, 1996). However, due to the acoustic properties of the intrauterine environment, this auditory experience is severely limited, to almost exclusively low-frequency sounds (Gerhardt and Abrams, 1996; Griffiths et al., 1994; Hepper and Shahidullah, 1994).

5.5.2 *A hypothesis for how degraded prenatal hearing may be adaptive*

The auditory experience of a fetus being confined to a predominantly lower frequency acoustic environment might be adaptive in a manner akin to the case of initially poor acuity in vision. Here, brief snippets of low-pass filtered auditory stimuli carrying a limited amount of sensory information provide a drive for gathering auditory information over extended time intervals. This drive would eventuate in the instantiation of neural mechanisms capable of supporting such integration of speech signals over larger temporal intervals. Given these changes in auditory information processing capacity, the HIA hypothesis would predict that commencing auditory experience with initially degraded inputs, devoid of higher frequencies, will result in extended temporal integration mechanisms and robust auditory analyses for tasks known to be especially reliant on it, such as emotion recognition or the analysis of other prosodic content (Ross et al., 1973; Snel and Cullen, 2013). To the contrary, having access to high auditory frequencies right from the outset would be expected to preclude the development of such mechanisms.

5.5.3 *Support from computational simulations for degraded prenatal hearing being adaptive*

Akin to the case of spatial frequency sensitivity, we had engaged in computational simulations, detailed in Vogelsang et al. (2023), in which we used a convolutional neural network (Dai et al., 2017) to probe the hypothesis of degraded inputs during prenatal hearing being adaptive for the acquisition of later auditory proficiencies. Specifically, we trained our network using several different regimens (one inspired by biological development; three others being non-biomimetic controls), on the temporally-extended task of emotion classification (Dupuis and Pichora-Fuller, 2010). The results from these simulations revealed that training with an auditory trajectory approximately recapitulating that of a neurotypical infant in the prenatal to postnatal period (i.e., transitioning from low-frequency to full-frequency inputs) resulted in temporally-extended receptive field structures in the first convolutional layer (analogous to the extended spatial receptive fields in the case of visual acuity) (see Figure 5.3C). Specifically, the receptive field structures were closer to those of a network trained exclusively on low frequencies than to those of a network trained exclusively on high frequencies (see Figures 5.3A-C). Furthermore, the developmentally-inspired model yielded the best subsequent generalization performance for emotion recognition on full-frequency and low-pass filtered speech snippets (see Figure 5.3D). These results reveal a remarkable correspondence between the visual and auditory modalities. They confirm the AID hypothesis' predictions that commencing auditory development with degraded, approximately low-pass filtered stimuli benefits later auditory proficiency, and also point to a potentially domain-general phenomenon.

5.5.4 *Empirical validation from prematurely born infants*

Interestingly, the computationally derived results, suggestive of the detrimental consequences of commencing auditory development with full-frequency inputs right from the start, instead of featuring initially low-pass filtered ones, are corroborated by data from prematurely born infants. Such individuals, whose experience with exclusively low-frequency sounds is artificially cut short, and who are almost immediately immersed into an environment teeming with high frequencies, have been shown to later exhibit impairments in the processing of the low-frequency structure of sounds, with resulting performance detriments on prosodic processing and emotion recognition, despite having near-normal punctate acoustic detection thresholds (Amin et

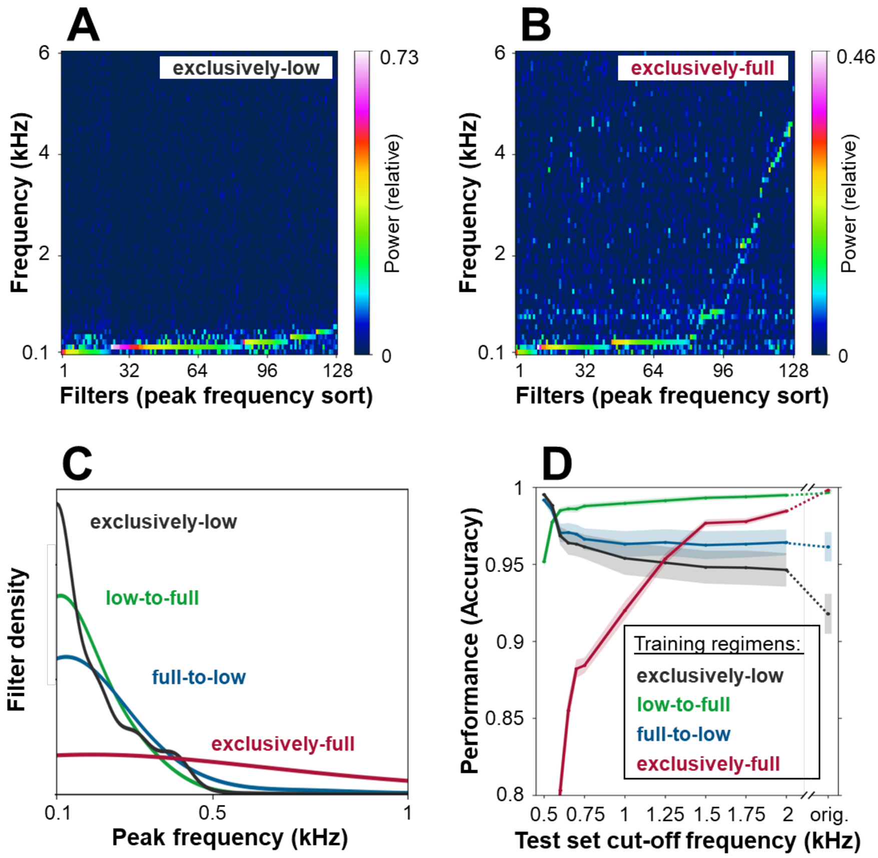


Figure 5.3: **A&B.** Spectral distribution of first-layer receptive fields in the network trained on exclusively low-frequency (A) and exclusively full-frequency (B) inputs. Colors code for normalized power. The results clearly show that exclusively-low training induces auditory receptive fields tuned to lower frequency content, relative to training on exclusively-full frequencies **C.** Kernel density estimation plot of the distribution of filters' peak frequencies for all networks, illustrating that training with the biomimetic regimen ('low-to-full' training) results in receptive field structure more similar to exclusively-low than to exclusively-full training. **D.** Mean and standard error of 10-fold cross-validated emotion recognition performances on full-frequency and various low-frequency test sets, demonstrating the generalization benefits of the low-to-full regimen in terms of performance on emotion recognition. These plots are adapted from Vogelsang et al. (2023).

al., 2015; Gonzalez-Gomez and Nazzi, 2012; Ragó et al., 2014). Taken together, the computational and empirical studies presented above make a compelling case for the adaptiveness of the degraded stimulus quality that is part of prenatal hearing.

5.6 POTENTIAL BUTTERFLY EFFECTS BEYOND LOW-TO-HIGH SPATIAL OR TEMPORAL FREQUENCY PROGRESSIONS

In addition to the developmental progression of low-to-high spatial frequencies in the visual domain and low-to-high temporal frequencies in the auditory domain, several other dimensions of perceptual development bear great promise for featuring additional 'butterfly effects'. While their empirical and/or computational validation is yet to come by, we present hypotheses of their potential adaptiveness for some of these dimensions below.

5.6.1 *Color sensitivity*

Studies on infant color vision, dating back to over a century ago (reviewed in Bornstein, 2006), strongly indicate that newborns are functionally color-deficient, in all three cone types (Adams and Courage, 2002; Suttle et al., 2002). These deficiencies are marked before eight weeks of age and gradually diminish by around 4-5 months of age (Kellman and Arterberry, 2007). How, if at all, might initial limitations in color sensitivity serve an adaptive role for later visual performance? We can draw upon the above work on acuity development to formulate a tentative answer. Analogously to immediate access to high spatial frequencies appearing to have the unfortunate consequence of constraining receptive fields to encode only local details, at the expense of extended spatial structures, the AID hypothesis would posit that, in the color domain, rich color information at the outset of vision may induce the developing image representations to strongly incorporate color cues, making the system overly reliant on such cues, and brittle when confronted with chromatic changes. Impoverished color information, as experienced by a typically developing infant, may help avoid this inducement. It remains to be seen whether this might, in part, explain our remarkable ability to recognize objects in grayscale images without much loss of accuracy (Biederman and Ju, 1988; Rossion and Pourtois, 2004).

5.6.2 *Joint acuity and color progressions in the visual domain*

While we have, thus far, discussed different developmental progressions independently, in real biological systems, these progressions are joint ones. For instance, both visual acuity and color sensitivity improve with age, albeit at different rates. Incorporating this joint development bears the potential of providing additional insights into visual function. One possibility could be the ability to explain a prominent organizing principle of response properties observed in the mammalian visual pathway. Several studies (e.g., Livingstone and Hubel, 1988; Van Essen et al., 1992) have reported that cells in the visual pathway can be broadly segregated into two groups: The magnocellular group exhibiting low-spatial frequency selectivity and low chromatic sensitivity; and the parvocellular group exhibiting high spatial frequency selectivity and high chromatic tuning. While this distinction is widely accepted, its genesis is unknown. The joint progression of color and acuity might provide a potential resolution: The start of visual experience is accompanied by low acuity and poor color sensitivity; hence, the cell response properties that emerge at this time could come to jointly encode these two attributes. But, as development progresses, higher acuity and richer color information become available, and the later developing cells may jointly encode these properties. Hence, the joint coding of low spatial frequency and low color information in some units, and high spatial frequency and high color sensitivity in others, could potentially be an outcome of the co-occurrence of these attributes at different time points during development.

5.6.3 *Language*

The applicability of the AID hypothesis might not only generalize across several dimensions of sensory development but could even extend into more cognitive ones, such as language acquisition. Specifically, the analysis of labels supplied by caregivers for visual entities in a child's environment reveals that the indicated categories progress from basic to subordinate: the parent of an infant is likely to point to a German shepherd, or a Scottish terrier, or a dachshund, and label them all as 'dogs'; only later would the subordinate distinctions be explicated (Callanan, 1985). Linguistic labels experienced by a young human learner therefore follow a coarse-to-fine progression. One possibility in which incorporation of such category hierarchy may impact the acquisition of visual classes, and the ability to learn visual features that can capture that hierarchy, is that the initial phase of training with basic-level labels may induce the system to identify characteristics that are common across the sub-categories (Gelman and Heyman, 1999; Keates and Graham, 2008; Lupyan, 2008; Lupyan et al., 2007; Waxman, 1999; Waxman and Booth, 2001, 2003; Waxman and Hall, 1993; Waxman and Markow, 1995). The immediate requirement to learn sub-categories may, on the other hand, build in a bias towards features that distinguish between the subclasses at the expense of detecting commonalities (Dewar and Xu, 2007, 2009; Feigenson and Halberda, 2008; Ferguson et al., 2015; Keates and Graham, 2008; Landau and Shipley, 2001; Scott and Monesson, 2009; Waxman and Braun, 2005; Xu, 2002; Xu et al., 2005; Zosh and Feigenson, 2009). This would render it difficult to subsequently group the sub-categories into a larger equivalence class. Further, starting with basic level categories and setting up corresponding representations may allow for the acquisition of subcategories with relatively few exemplars for each subclass, just enough to augment the basic level representation. Thus, the 'biomimetic' label progression may help set up basic visual representations, which can then, upon the availability of subordinate labels, enable rapid acquisition of hierarchical class structure. This would yield visual representations that conform to a natural hierarchical category structure and might be one of the origins of why we exhibit a remarkably robust ability to group diverse exemplars into basic categories (such as categorizing Dalmatians, Alsatians, and chihuahuas all as 'dogs').

While systematic empirical and/or computational investigations of the potential 'butterfly effects' presented above are yet to come by, they are illustrative of the potential of broadening the AID hypothesis to additional key dimensions of perceptual development.

5.7 DISCUSSION

In this paper, we have reviewed recent findings from empirical and computational studies, from our lab and others, that have provided compelling evidence for the proposal that initial degradations in early sensory experience are likely to be scaffolds, rather than hurdles, for the acquisition of later perceptual skills.

There are a few important caveats to be pointed out. First, the computational results presented were derived from simulations with convolutional neural networks as computational model systems. The fealty to these networks is not intended to assert that they are perfect biological models. However, they can meaningfully complement the presented empirical studies by allowing for a more direct examination of the impact of specific training regimens while being rough proxies of the biological system. Similarly, the biomimetic training regimens are designed to capture the basic structure of developmental progressions but not all the details. This is admittedly a limitation, but a reasoned one. A rough approximation may be adequate to validate the AID hypothesis, and trying to achieve hyper-fidelity is likely to be subject to diminishing returns – some of the fine-grained details of developmental progressions may be specific to 'hardware' and may not contribute significantly to our understanding of how developmental change impacts skill acquisition.

From the perspective of artificial intelligence, as pointed out in Katzhendler and Weinshall (2019), mimicking the specific developmental staging from limited to proficient may not be favored over an artificially-engineered approach, in which the ordering of degraded and non-degraded stimuli would be a randomly mixed one, lacking implementational feasibility in the developing biological system. Further, our proposal bears conceptual linkages with the notion of curriculum learning (Bengio et al., 2009; Hacohen and Weinshall, 2019), whose overarching idea is to "organize the [training] examples into gradually more complex ones" (Bengio et al., 2009) and has been suggested to benefit the speed of training convergence and generalization. Our proposal is a particular variant of this general strategy as, of the many possible curricula, we investigate those that transpire developmentally. Such sequencing may not necessarily be consistent with the lower to higher complexity scheme of conventional curriculum learning. Hence, while there are important linkages to be drawn between developmental progressions and typical curricular sequences, the former merit further independent study in their own right to reveal whether the choreography of biological development is especially beneficial for learning outcomes.

Finally, it is worth pointing out that the finding that certain aspects of development appear to prove beneficial for subsequent performance does not prove that they are intended to produce those effects. The benefits could nevertheless be epiphenomenal. Further, while we have discussed butterfly effects across several perceptual dimensions, it is important to note that some kinds of perceptual proficiencies are likely to not find their roots in initial developmental degradations. To the contrary, the early availability of some of these proficiencies might, in fact, be causally responsible for supporting the development of other proficiencies. A differential examination of the developmental profiles of these factors might, in the future, thus, provide further insight into their interactions.

5.8 conclusion & implications

Notwithstanding the above caveats, we believe that the reviewed work on the AID hypothesis has far-reaching implications for the domains of developmental science, artificial intelligence, and clinical practice. First, from the developmental science perspective, a comprehensive understanding of perceptual development requires that descriptions of developmental stages be augmented with hypotheses about their functional significance. Findings from the presented work are providing such hypotheses, and thereby an account for why normal perceptual development follows its stereotypical trajectories. The converging empirical and computational evidence presented thus far already makes a compelling case for the applicability of the HIA proposals to the domains of visual acuity and prenatal hearing, with the generalization to additional dimensions of sensory development yet to be explored.

For the domain of artificial intelligence, despite the stated limitations, biomimetic regimens can, in certain scenarios, already now guide the creation of training strategies for improving deep networks, as shown, for instance, in the context of visual acuity-inspired curricula for medical image classification (Basu et al., 2022; Chen et al., 2023). More generally, the presented work illustrates the potential benefits of incorporating aspects of human development into the design of training routines for deep networks. The latter has more broadly also been elaborated on in Zaadnoordijk et al. (2022), in which the researchers highlight three insights from infant development that could help further improve unsupervised machine learning, to wit: an infant's guided and constrained information processing, the presentation of diverse and multimodal inputs, and the shaping of inputs by development and active learning. In addition, the current work also serves as an example of how machine learning systems can be used to investigate biological development.

Finally, from the applied clinical perspective, millions of children suffer from conditions such as premature births, deafness, and congenital cataracts that alter their early sensory experience. It is important to have accurate prognoses for how atypical early experiences are going to impact later proficiencies, so that appropriate rehabilitative interventions can be designed. The presented work can provide the principles for formulating such prognoses and interventions. For instance, in the case of premature births, neonatal ICUs do not typically allow newborns to be exposed to primarily low-frequency sounds (Lahav, 2015). To the contrary, if any auditory interventions are applied, they are typically focused on the presence of musical sounds (De Almeida et al., 2020; Loewy et al., 2013; Lordier et al., 2019) or the induction of complete silence (Altuncu et al., 2009; Milette, 2010). One possibility to improve the current clinical standard would be to filter environmental sounds with an approximation of the acoustic properties of the intrauterine environment. Similarly, in the case of late restoration of congenital blindness, instead of exposing newly sighted children to high-quality visual inputs immediately following surgery, it might prove beneficial to artificially limit the stimulus quality initially and gradually increase the complexity of environmental stimuli thereafter. While, as of now, many of these rehabilitative measures are just speculation; considering that the presented work was seeded by our meeting RK, it would be deeply gratifying to see it having implications for clinical practice and a meaningful impact on the lives of many other children like him.

REFERENCES

Adams, Russell J and Mary L Courage (2002). "A psychophysical test of the early maturation of infants' mid-and long-wavelength retinal cones." In: *Infant Behavior and Development* 25.2, pp. 247–254.

Altuncu, E, I Akman, S Kulekci, F Akdas, Hülya Bilgen, and E Ozek (2009). "Noise levels in neonatal intensive care unit and use of sound absorbing panel in the isolette." In: *International journal of pediatric otorhinolaryngology* 73.7, pp. 951–953.

Amin, Sanjiv B, Mark Orlando, Christy Monczynski, and Kim Tillery (2015). "Central auditory processing disorder profile in premature and term infants." In: *American journal of perinatology* 32.04, pp. 399–404.

Anandan, Padmanabhan (1989). "A computational framework and an algorithm for the measurement of visual motion." In: *International Journal of Computer Vision* 2.3, pp. 283–310.

Arcaro, Michael J, Peter F Schade, Justin L Vincent, Carlos R Ponce, and Margaret S Livingstone (2017). "Seeing faces is necessary for face-domain formation." In: *Nature neuroscience* 20.10, pp. 1404–1412.

Banks, Martin S and Patrick J Bennett (1988). "Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates." In: *JOSA A* 5.12, pp. 2059–2079.

Banks, Martin S and James A Crowell (1993). "Front-end limitations to infant spatial vision: Examination of two analyses." In: *Early visual development: Normal and abnormal*, pp. 91–116.

Basu, Soumen, Mayank Gupta, Pratyaksha Rana, Pankaj Gupta, and Chetan Arora (2022). "Surpassing the human accuracy: detecting gallbladder cancer from USG images with curriculum learning." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20886–20896.

Bengio, Yoshua, Jérôme Louradour, Ronan Collobert, and Jason Weston (2009). "Curriculum learning." In: *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48.

Biederman, Irving and Ginny Ju (1988). "Surface versus edge-based determinants of visual recognition." In: *Cognitive psychology* 20.1, pp. 38–64.

Black, Michael J and Paul Anandan (1996). "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields." In: *Computer vision and image understanding* 63.1, pp. 75–104.

Boas, Judith AR, RL Ramsey, AH Riesen, and JP Walker (1969). "Absence of change in some measures of cortical morphology in dark-reared adult rats." In: *Psychonomic Science* 15.5, pp. 251–252.

Bornstein, Marc H (2006). "Hue categorization and color naming physics To sensation to perception." In: *Progress in Colour Studies: Volume II. Psychological aspects*, p. 35.

Cadena, Santiago A, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, and Alexander S Ecker (2019). "Deep convolutional models improve predictions of macaque V1 responses to natural images." In: *PLoS computational biology* 15.4, e1006897.

Callanan, Maureen A (1985). "How parents label objects for young children: The role of input in the acquisition of category hierarchies." In: *Child development*, pp. 508–523.

Chen, Xiong, Guochang You, Qinchang Chen, Xiangxiang Zhang, Na Wang, Xuehua He, Liling Zhu, Zhouzhou Li, Chen Liu, Shixiang Yao, et al. (2023). "Development and evaluation of an artificial intelligence system for children intussusception diagnosis using ultrasound images." In: *Iscience* 26.4.

Courage, Mary L and Russell J Adams (1990). "Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples." In: *Optometry and vision science* 67.9, pp. 713–718.

Dai, Wei, Chia Dai, Shuhui Qu, Juncheng Li, and Samarjit Das (2017). "Very deep convolutional neural networks for raw waveforms." In: *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, pp. 421–425.

Dakin, Steven C and Andrew M Herbert (1998). "The spatial region of integration for visual symmetry detection." In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265.1397, pp. 659–664.

Daw, Nigel W (2014). *Visual development*. Vol. 14. Springer New York, NY.

De Almeida, Joana Sa, Lara Lordier, Benjamin Zollinger, Nicolas Kunz, Matteo Bastiani, Laura Gui, Alexandra Adam-Darque, Cristina Borradori-Tolsa, François Lazeyras, and Petra S Hüppi (2020). "Music enhances structural maturation of emotional processing neural pathways in very preterm infants." In: *Neuroimage* 207, p. 116391.

De Heering, Adélaïde and Daphne Maurer (2014). "Face memory deficits in patients deprived of early visual input by bilateral congenital cataracts." In: *Developmental Psychobiology* 56.1, pp. 96–108.

De Schonen, Scania and Eric Mathivet (1989). "First come, first served: A scenario about the development of hemispheric specialization in face recognition during infancy." In: *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*.

Dewar, Kathryn and Fei Xu (2007). "Do 9-month-old infants expect distinct words to refer to kinds?" In: *Developmental psychology* 43.5, p. 1227.

Dewar, Kathryn and Fei Xu (2009). "Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants." In: *Psychological Science* 20.2, pp. 252–257.

Dobkins, Karen R, Barry Lia, and Davida Y Teller (1997). "Infant color vision: Temporal contrast sensitivity functions for chromatic (red/-green) stimuli in 3-month-olds." In: *Vision Research* 37.19, pp. 2699–2716.

Dobson, Velma and Davida Y Teller (1978). "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies." In: *Vision research* 18.11, pp. 1469–1483.

Dominguez, Melissa and Robert A Jacobs (2003). "Developmental constraints aid the acquisition of binocular disparity sensitivities." In: *Neural Computation* 15.1, pp. 161–182.

Dupuis, Kate and M Kathleen Pichora-Fuller (2010). *Toronto emotional speech set (TESS)*.

Elman, Jeffrey L (1993). "Learning and development in neural networks: The importance of starting small." In: *Cognition* 48.1, pp. 71–99.

Feigenson, Lisa and Justin Halberda (2008). "Conceptual knowledge increases infants' memory capacity." In: *Proceedings of the National Academy of Sciences* 105.29, pp. 9926–9930.

Ferguson, Brock, Mélanie Havy, and Sandra R Waxman (2015). "The precision of 12-month-old infants' link between language and categorization predicts vocabulary size at 12 and 18 months." In: *Frontiers in psychology* 6, p. 1319.

Field, David J, Anthony Hayes, and Robert F Hess (1993). "Contour integration by the human visual system: evidence for a local "association field"." In: *Vision research* 33.2, pp. 173–193.

Geldart, Sybil, Catherine J Mondloch, Daphne Maurer, Scania De Schonen, and Henry P Brent (2002). "The effect of early visual deprivation on the development of face processing." In: *Developmental Science* 5.4, pp. 490–501.

Gelman, Susan A and Gail D Heyman (1999). "Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories." In: *Psychological Science* 10.6, pp. 489–493.

Gerhardt, Kenneth J and Robert M Abrams (1996). "Fetal hearing: characterization of the stimulus and response." In: *Seminars in perinatology*. Vol. 20. 1. Elsevier, pp. 11–20.

Goffaux, Valerie, Barbara Hault, Caroline Michel, Quoc C Vuong, and Bruno Rossion (2005). "The respective role of low and high spatial frequencies in supporting configural and featural processing of faces." In: *Perception* 34.1, pp. 77–86.

Gonzalez-Gomez, Nayeli and Thierry Nazzi (2012). "Phonotactic acquisition in healthy preterm infants." In: *Developmental science* 15.6, pp. 885–894.

Griffiths, Scott K, WS Brown Jr, Kenneth J Gerhardt, Robert M Abrams, and Richard J Morris (1994). "The perception of speech sounds recorded within the uterus of a pregnant sheep." In: *The Journal of the Acoustical Society of America* 96.4, pp. 2055–2063.

Hacohen, Guy and Daphna Weinshall (2019). "On the power of curriculum learning in training deep networks." In: *International conference on machine learning*. PMLR, pp. 2535–2544.

Hendrickson, Anita and Ronald Boothe (1976). "Morphology of the retina and dorsal lateral geniculate nucleus in dark-reared monkeys (Macaca nemestrina)." In: *Vision research* 16.5, 517–IN5.

Hepper, Peter G and B Sara Shahidullah (1994). "The development of fetal hearing." In: *Fetal and Maternal Medicine Review* 6.3, pp. 167–179.

Ince, Robin AA, Nicola J Van Rijsbergen, Gregor Thut, Guillaume A Rousselet, Joachim Gross, Stefano Panzeri, and Philippe G Schyns (2015). "Tracing the flow of perceptual features in an algorithmic brain network." In: *Scientific reports* 5.1, p. 17681.

Jacobs, Deborah S and Colin Blakemore (1988). "Factors limiting the postnatal development of visual acuity in the monkey." In: *Vision research* 28.8, pp. 947–958.

Jang, Hojin and Frank Tong (2021). "Convolutional neural networks trained with a developmental sequence of blurry to clear images reveal core differences between face and object processing." In: *Journal of vision* 21.12, pp. 6–6.

Jinsi, Omisa, Margaret M Henderson, and Michael J Tarr (2023). "Early experience with low-pass filtered images facilitates visual category learning in a neural network model." In: *Plos one* 18.1, e0280145.

Katzhendler, Gal and Daphna Weinshall (2019). "Potential upside of high initial visual acuity?" In: *Proceedings of the National Academy of Sciences* 116.38, pp. 18765–18766.

Keates, Jean and Susan A Graham (2008). "Category markers or attributes: why do labels guide infants' inductive inferences?" In: *Psychological Science* 19.12, pp. 1287–1293.

Kellman, Philip J and Martha E Arterberry (2007). "Infant visual perception." In: *Handbook of child psychology* 2.

Kiorpes, Lynne (2016). "The puzzle of visual development: behavior and neural limits." In: *Journal of Neuroscience* 36.45, pp. 11384–11393.

Kiorpes, Lynne and J Anthony Movshon (2004). "Neural limitations on visual development in primates." In: *The visual neurosciences* 1, pp. 159–173.

Kong, Nathan CL, Eshed Margalit, Justin L Gardner, and Anthony M Norcia (2022). "Increasing neural network robustness improves match to macaque V1 eigenspectrum, spatial frequency preference and predictivity." In: *PLOS Computational Biology* 18.1, e1009739.

Kriegeskorte, Nikolaus (2015). "Deep neural networks: a new framework for modeling biological vision and brain information processing." In: *Annual review of vision science* 1, pp. 417–446.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems* 25.

Kwon, MiYoung, Rong Liu, and Lillian Chien (2016). "Compensation for blur requires increase in field of view and viewing time." In: *PLoS One* 11.9, e0162711.

Lahav, Amir (2015). "Questionable sound exposure outside of the womb: frequency analysis of environmental noise in the neonatal intensive care unit." In: *Acta paediatrica* 104.1, e14–e19.

Landau, Barbara and Elizabeth Shipley (2001). "Labelling patterns and object naming." In: *Developmental science* 4.1, pp. 109–118.

Lenassi, Eva, Katarina Likar, Branka Stirn-Kranjc, and Jelka Brecelj (2008). "VEP maturation and visual acuity in infants and preschool children." In: *Documenta Ophthalmologica* 117, pp. 111–120.

Lindsay, Grace W (2021). "Convolutional neural networks as a model of the visual system: Past, present, and future." In: *Journal of cognitive neuroscience* 33.10, pp. 2017–2031.

Livingstone, Margaret and David Hubel (1988). "Segregation of form, color, movement, and depth: anatomy, physiology, and perception." In: *Science* 240.4853, pp. 740–749.

Loewy, Joanne, Kristen Stewart, Ann-Marie Dassler, Aimee Telsey, and Peter Homel (2013). "The effects of music therapy on vital signs, feeding, and sleep in premature infants." In: *Pediatrics* 131.5, pp. 902–918.

Lordier, Lara, Djalel-Eddine Meskaldji, Frédéric Grouiller, Marie P Pittet, Andreas Vollenweider, Lana Vasung, Cristina Borradori-Tolsa, François Lazeyras, Didier Grandjean, Dimitri Van De Ville, et al. (2019). "Music in premature infants enhances high-level cognitive brain networks." In: *Proceedings of the National Academy of Sciences* 116.24, pp. 12103–12108.

Lorenz, Edward N. (1963). "Section of planetary sciences: the predictability of hydrodynamic flow." In: *Transactions of the New York Academy of Sciences* 25.4 Series II, pp. 409–432.

Lupyan, Gary (2008). "From chair to" chair": a representational shift account of object labeling effects on memory." In: *Journal of Experimental Psychology: General* 137.2, p. 348.

Lupyan, Gary, David H Rakison, and James L McClelland (2007). "Language is not just for talking: Redundant labels facilitate learning of novel categories." In: *Psychological science* 18.12, pp. 1077–1083.

Mémin, Etienne and Patrick Pérez (2002). "Hierarchical estimation and segmentation of dense motion fields." In: *International Journal of Computer Vision* 46, pp. 129–155.

Milette, Isabelle (2010). "Decreasing noise level in our NICU: the impact of a noise awareness educational program." In: *Advances in Neonatal Care* 10.6, pp. 343–351.

Nelson, Charles A (2001). "The development and neural bases of face recognition." In: *Infant and Child Development: An International Journal of Research and Practice* 10.1-2, pp. 3–18.

Newport, Elissa L (1988). "Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language." In: *Language sciences* 10.1, pp. 147–172.

Newsome, William T and Edmond B Pare (1988). "A selective impairment of motion perception following lesions of the middle temporal visual area (MT)." In: *Journal of Neuroscience* 8.6, pp. 2201–2211.

Ng, Hong-Wei and Stefan Winkler (2014). "A data-driven approach to cleaning large face datasets." In: *2014 IEEE international conference on image processing (ICIP)*. IEEE, pp. 343–347.

Peterson, Mary A and Gillian Rhodes (2003). *Perception of faces, objects, and scenes: Analytic and holistic processes*. Oxford University Press.

Putzar, Lisa, Kirsten Hötting, and Brigitte Röder (2010). "Early visual deprivation affects the development of face recognition and of audio-visual speech perception." In: *Restorative neurology and neuroscience* 28.2, pp. 251–257.

Ragó, Anett, Ferenc Honbolygó, Zsófia Róna, Anna Beke, and Valéria Csépe (2014). "Effect of maturation on suprasegmental speech processing in full-and preterm infants: A mismatch negativity study." In: *Research in developmental disabilities* 35.1, pp. 192–202.

Richler, Jennifer J and Isabel Gauthier (2014). "A meta-analysis and review of holistic face processing." In: *Psychological bulletin* 140.5, p. 1281.

Rivolta, Davide (2014). "Cognitive and Neural Aspects of Face Processing." In: *Prosopagnosia: When all faces look the same*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 19–40. ISBN: 978-3-642-40784-0.

Röder, Brigitte, Pia Ley, Bhamy H Shenoy, Ramesh Kekunnaya, and Davide Bottari (2013). "Sensitive periods for the functional specialization of the neural system for human face processing." In: *Proceedings of the National Academy of Sciences* 110.42, pp. 16760–16765.

Ross, Mark, Robert J Duffy, Harry S Cooker, and Russell L Sargeant (1973). "Contribution of the lower audible frequencies to the recognition of emotions." In: *American Annals of the Deaf*, pp. 37–42.

Rossion, Bruno and Gilles Pourtois (2004). "Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition." In: *Perception* 33.2, pp. 217–236.

Schrimpf, Martin, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo (2020). "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?" In: *bioRxiv*.

Scott, Lisa S and Alexandra Monesson (2009). "The origin of biases in face perception." In: *Psychological Science* 20.6, pp. 676–680.

Sizintsev, Mikhail and Richard P Wildes (2010). "Coarse-to-fine stereo vision with accurate 3D boundaries." In: *Image and Vision Computing* 28.3, pp. 352–366.

Skoczenski, Ann M and Anthony M Norcia (1998). "Neural noise limitations on infant visual sensitivity." In: *Nature* 391.6668, pp. 697–700.

Smith, Fraser W and Philippe G Schyns (2009). "Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances." In: *Psychological Science* 20.10, pp. 1202–1208.

Snel, John and Charlie Cullen (2013). "Judging emotion from low-pass filtered naturalistic emotional speech." In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, pp. 336–342.

Storrs, Katherine R, Tim C Kietzmann, Alexander Walther, Johannes Mehrer, and Nikolaus Kriegeskorte (2021). "Diverse deep neural networks all predict human inferior temporal cortex well, after training and fitting." In: *Journal of cognitive neuroscience* 33.10, pp. 2044–2064.

Suttle, Catherine M, Martin S Banks, and Erich W Graf (2002). "FPL and sweep VEP to tritan stimuli in young human infants." In: *Vision Research* 42.26, pp. 2879–2891.

Taubert, Jessica, Deborah Apthorp, David Aagten-Murphy, and David Alais (2011). "The role of holistic processing in face perception: Evidence from the face inversion effect." In: *Vision research* 51.11, pp. 1273–1278.

Turkewitz, Gerald and Patricia A Kenny (1982). "Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement." In: *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 15.4, pp. 357–368.

Van Essen, David C, Charles H Anderson, and Daniel J Felleman (1992). "Information processing in the primate visual system: an integrated systems perspective." In: *Science* 255.5043, pp. 419–423.

Vogelsang, Lukas, Sharon Gilad-Gutnick, Sidney Diamond, Albert Yonas, and Pawan Sinha (2019). "Response to Katzhendler and Weinshall: Initial visual degradation during development may be adaptive." In: *Proceedings of the National Academy of Sciences* 116.38, pp. 18767–18768.

Vogelsang, Lukas, Sharon Gilad-Gutnick, Evan Ehrenberg, Albert Yonas, Sidney Diamond, Richard Held, and Pawan Sinha (2018). "Potential downside of high initial visual acuity." In: *Proceedings of the National Academy of Sciences* 115.44, pp. 11333–11338.

Vogelsang, Marin, Lukas Vogelsang, Sidney Diamond, and Pawan Sinha (2023). "Prenatal auditory experience and its sequelae." In: *Developmental Science* 26.1, e13278.

Waxman, Sandra R (1999). "Specifying the scope of 13-month-olds' expectations for novel words." In: *Cognition* 70.3, B35–B50.

Waxman, Sandra R and Amy E Booth (2001). "Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives." In: *Cognitive psychology* 43.3, pp. 217–242.

Waxman, Sandra R and Amy E Booth (2003). "The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds." In: *Developmental Science* 6.2, pp. 128–135.

Waxman, Sandra R and Irena Braun (2005). "Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants." In: *Cognition* 95.3, B59–B68.

Waxman, Sandra R and D Geoffrey Hall (1993). "The development of a linkage between count nouns and object categories: Evidence from fifteen-to twenty-one-month-old infants." In: *Child development* 64.4, pp. 1224–1241.

Waxman, Sandra R and Dana B Markow (1995). "Words as invitations to form categories: Evidence from 12-to 13-month-old infants." In: *Cognitive psychology* 29.3, pp. 257–302.

Wilson, HR (1993). "Theories of infant visual development." In: *Early visual development: Normal and abnormal*, pp. 560–572.

Xu, Fei (2002). "The role of language in acquiring object kind concepts in infancy." In: *Cognition* 85.3, pp. 223–250.

Xu, Fei, Melissa Cote, and Allison Baker (2005). "Labeling guides object individuation in 12-month-old infants." In: *Psychological Science* 16.5, pp. 372–377.

Young, Andrew W, Deborah Hellawell, and Dennis C Hay (1987). "Configurational information in face perception." In: *Perception* 16.6, pp. 747–759.

Yuodelis, Cristine and Anita Hendrickson (1986). "A qualitative and quantitative analysis of the human fovea during development." In: *Vision research* 26.6, pp. 847–855.

Zaadnoordijk, Lorijn, Tarek R Besold, and Rhodri Cusack (2022). "Lessons from infant learning for unsupervised machine learning." In: *Nature Machine Intelligence* 4.6, pp. 510–520.

Zosh, Jennifer M and Lisa Feigenson (2009). "Beyond 'what'and 'how many': capacity, complexity and resolution of infants' object representations." In: *The origins of object knowledge*, pp. 25–51.

6 DEVELOPMENT OF VISUAL MEMORY CAPACITY FOLLOWING EARLY-ONSET AND EXTENDED BLINDNESS

Content from

Gupta, P., Shah, P., Gilad-Gutnick, S., **Vogelsang, M.**, Vogelsang, L., Tiwari, K., Gandhi, T., Ganesh, S., & Sinha, P. (2022). "Development of visual memory capacity following early-onset and extended blindness". **Published** in Psychological Science, 33(6), 847-858.

6.1 ABSTRACT

It is unknown whether visual memory capacity can develop if onset of pattern vision is delayed for several years following birth. We had an opportunity to address this question through our work with an unusual population of 12 congenitally blind individuals ranging in age from 8 to 22 years. After providing them with sight surgery, we longitudinally evaluated their visual memory capacity using an image-memorization task. Our findings revealed poor visual memory capacity soon after surgery but significant improvement in subsequent months. Although there may be limits to this improvement, performance 1 year after surgery was found to be comparable with that of control participants with matched visual acuity. These findings provide evidence for plasticity of visual memory mechanisms into late childhood but do not rule out vulnerability to early deprivation. Our computational simulations suggest that a potential mechanism to account for changes in memory performance may be progressive representational elaboration in image encoding.

6.2 KEYWORDS

cognitive development, visual memory, impact of early visual deprivation, late sight onset

6.3 STATEMENT OF RELEVANCE

Humans exhibit impressive visual memory from early in life. What are the roots of this proficiency? Specifically, does its development depend on early visual experience? These fundamental questions have proven difficult to answer given the challenges of working with neonates and the ethical impossibility of limiting their visual experience. We had the opportunity to address this issue through our work with an unusual group of children, those who had been born blind and did not receive treatment for several years. We provided them with sight surgeries and then longitudinally studied the development of visual memory capacity. Our results revealed a steady improvement in their memory performance and indicated the feasibility of neural change even late in childhood. We also describe computational simulations that provide hints about the possible nature of neural changes underlying the observed improvements in visual memory performance.

6.4 INTRODUCTION

Given the crucial role visual memory plays in enabling many aspects of cognitive function, it is not surprising that visual memory develops very rapidly. Visual memory capacity increases significantly over the first year of life (Rose et al., 2001; Ross-sheehy et al., 2003) and continues to increase with age in childhood (Cowan et al., 2011; Riggs et al., 2006; Simmering, 2012; Walker et al., 1994). Children as young as 4 years of age spontaneously encode a high degree of visual detail. They exhibit high fidelity in their visual memory capacity over a large set of items not only for basic-level categories but also for unique details and information about the position and arrangement of parts (Ferrara et al., 2017). By adulthood, humans come to possess the ability to remember several thousand images, which manifests as high accuracy in immediate recognition tests (Brady et al., 2008; Madigan, 2014; Nickerson, 1965; Shepard, 1967; Standing, 1973; Standing et al., 1970). This impressive performance has even led some researchers to speculate that picture-memory capacity may be essentially unlimited (Yuille, 2014).

Notwithstanding evidence of excellent picture-memory capacity, several questions about the development of this ability remain open. One of these concerns the role of early visual experience: Can visual memory develop even if visual experience is precluded in the first several years of life? Past research on animals has shown that early visual deprivation can cause dramatic changes in neural organization that have a profound impact on many visual functions (Hubel and Wiesel, 1970). Human studies of recovery after congenital blindness have shown similar trends (Lewis and Maurer, 2005; Ostrovsky et al., 2006). Early visual deprivation is found to have profound consequences on the subsequent development of various low-level visual functions, such as acuity, contrast sensitivity, and motion coherence (Braddick et al., 2003; Cobb and MacDonald, 1978; Ellemberg et al., 2002; Kalia et al., 2014; Sinha and Held, 2012). Other studies of visual function after treatment of congenital blindness suggest that early deprivation also impairs high-level aspects of vision, such as object and face recognition (Fine et al., 2003; Gregory and Wallace, 1963; Sacks, 1995; Valvo et al., 1971; Von Senden, 1960). Even relatively short periods of deprivation ranging in duration from 2 to 6 months after birth have been shown to have significant detrimental consequences on face-recognition skills (Le Grand et al., 2001; Lewis and Maurer, 2005; Maurer et al., 2005, 2007). Given this evidence of compromised visual perception after late sight onset, and the suggestion from several studies (reviewed in Slotnick, 2004) that visual memory and visual perception recruit common neural substrates, it is possible that visual memory may be vulnerable to early deprivation in the same way that visual perception is. However, there are also plausible reasons to expect resilience of memory resources. A notable one is the possibility of memory being amodal, catering to and sustained by incoming information from multiple sensory modalities. In this perspective, visual deprivation leads to a diminution, rather than elimination, of input to the amodal memory store. Reports such as those of Wolbers et al. (2011) point to the biological feasibility of such a proposal. Taken together, the mixed nature of past results highlights the difficulty of definitively predicting how visual deprivation would impact the subsequent development of image memory. Accordingly, our goal here was to examine this question empirically by investigating whether the development of visual memory capacity is impacted by early and extended visual deprivation.

We have had an opportunity to address this question as part of a humanitarian and scientific initiative, Project Prakash (Sinha, 2013), that is intended to alleviate curable childhood blindness in the developing world while also addressing scientific questions about visual development. As part of this effort, we provide free surgical treatment to congenitally blind children and study changes in their visual skills and associated neural reorganization as the children gain visual experience. Working with this unusual population provides a glimpse into the impact of early visual deprivation on visual skill acquisition and the interplay between nature and nurture during that process.

In the present study, we followed the development of visual memory in 12 newly sighted Prakash patients. Participants performed an old/new task that involved memorizing a set of images and then completing a recall task in which they were asked to identify the memorized items from a larger collection that included previously seen images and novel distractors. The sets to be memorized differed in their cardinalities (i.e., the number of images to be memorized) and the semantic information that they contained (images of natural scenes vs. abstract paintings). The real and abstract sets were comparable in low-level image properties such as luminance and spatial frequency. We were thus able to examine the posttreatment development of visual memory capacity and, using the two types of stimulus sets that differed in their semantic content, also focus the investigation to study the potential emergence of meaning as a modulator of this capacity.

Previous studies that investigated whether meaning influences memory have typically done so with linguistic stimuli; relatively few studies have examined this issue purely in the visual domain. Notably, Bellhouse-King and Standing (2007) compared visual memory performance of concrete, regular abstract, and diverse abstract pictures. The investigators found that meaningfulness is not necessary for effective picture memory but does appear to facilitate it. They found that even quite meaningless visual designs can be recognized at well above chance level, provided that they are sufficiently distinctive so that they are not confused with each other. However, the abstract images used in the study were pictures of snowflakes (regular abstract) and groupings of simple geometrical figures such as triangles, rectangles, and circles (diverse abstract), so all patterns belonged to the same conceptual class (snowflakes) or shared the same small set of constituent elements.

Building on this past work, we sought to address the primary question of whether visual memory capacity could develop even after several years of congenital blindness and, secondarily, whether and to what extent image semantics contribute to memory capacity.

6.5 METHOD

6.5.1 *Participants*

We recruited two groups of participants. The first group comprised 12 newly sighted Prakash patients (six females) ranging in age from 8 to 22 years (M = 12.8 years). The size of the experimental group was determined by the number of children who met the inclusion criteria (treatable congenital profound visual deprivation) during the study period. The patients were tested prior to treatment for congenital cataracts and periodically up to a year afterward. The control group comprised 12 normally sighted schoolchildren ranging in age from 8 to 14 years (M = 10.8 years). This study was approved by the Massachusetts Institute of Technology Institutional Review Board (Committee on the Use of Humans as Experimental Subjects) and the Ethics Committee of Dr. Shroff's Charity Eye Hospital, New Delhi, our medical partner.

6.5.1.1 *Patient group*

All patients had been identified via our project's pediatric ophthalmic screening program in rural areas of India. All had dense bilateral cataracts since before 1 year of age. Assessment of congenitality of deprivation was based on parental reports, ophthalmic examination of the eyeball and cataract morphology, and the presence of nystagmus, which is known to be induced by profound visual impairment very early in life (Tusa et al., 1991). It should be noted that although these factors render an inference of congenitality highly probable, they are not absolutely definitive.

6.5.1.2 *Preoperative assessments*

We tested for light perception in all four quadrants and measured acuity using the Freiberg Visual Acuity Test (Bach et al., 1996). The patient information table (see Table 6.3 in the Supplemental Material available online) shows that all individuals were in the category of "Profound visual impairment (20/500–20/1000)" or "Light perception/projection (< 20/1000)," as per classification norms followed by the American Foundation for the Blind (2020). The anterior segment was evaluated with a slit lamp, and the type of cataract and any associated ocular pathology were noted. Given the patients' dense bilateral cataracts (which precluded fundus viewing via ophthalmoscopes), B-scan ultrasound was recorded before surgery in all cases to check for any posterior segment pathology.

6.5.1.3 *Intervention*

Keratometry and biometry of all children were performed under general anesthesia immediately preceding the surgery. All surgeries were performed by a single surgeon. The patients underwent a primary posterior capsulorhexis through the anterior route, and a foldable acrylic posterior chamber intraocular lens was implanted in the bag. The best refractive correction was prescribed to patients after their sutures were removed.

6.5.1.4 *Control group*

Participants in the control group were recruited from a school in New Delhi. They had normal or corrected-to-normal vision. These participants had no history of neurological or psychiatric illness and their socioeconomic status was matched with that of the patients. Informed assent was obtained from all participants, and consent was obtained from their parents.

6.5.2 *Stimuli*

For the stimulus sets, we compiled a database of 1,200 real-world and 1,200 abstract images. All images were cropped square and scaled to the same size (256 × 256 pixels). Real images depicted a variety of natural scenes including architecture, flora, fauna, vehicles, and people (Fig. 6.1a). Abstract images were nonrepresentational paintings drawn from several digital art archives (Fig. 6.1b). In order to determine whether the inherent discriminability of the images in the two sets was comparable, we computed the 2D correlation coefficients of

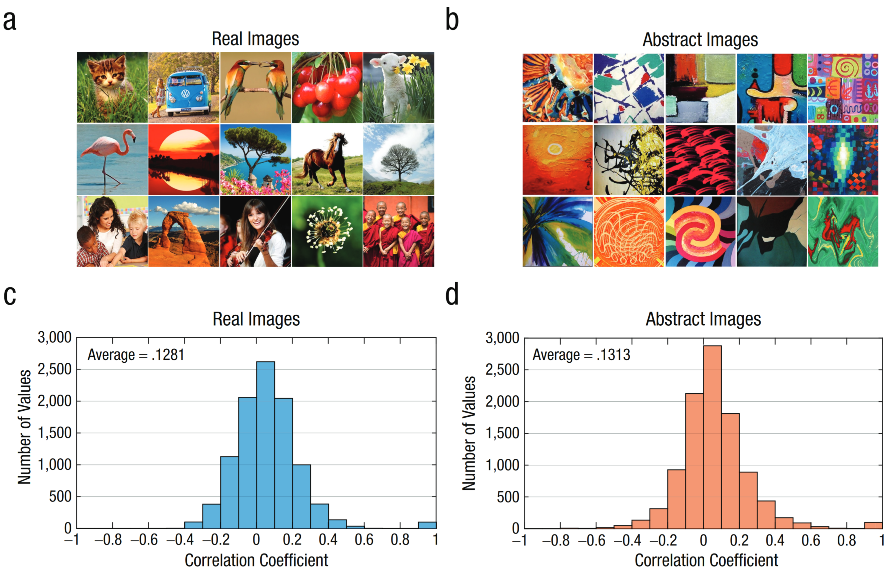


Figure 6.1: Example stimuli and stimulus characteristics. The top row shows sample (a) real-world scenes and (b) abstract paintings used in our study. The full stimulus set comprised 2,400 images (1,200 of each type). The distributions in the bottom row show 2D correlation coefficients for all pairwise comparisons across (c) 100 randomly selected real images and (d) 100 randomly selected abstract images.

all pairwise comparisons across 100 randomly selected real images and 100 randomly selected abstract images. We found that the two distributions were statistically indistinguishable (Figs. 6.1c and6.1d). An unpaired-samples t test found that the correlation between real and abstract images was not significant, $t(9898) = -0.9364$, $p = .3491$, $SD = 0.1626$, 95% confidence interval $= [-0.0095, 0.0033]$.

6.5.3 *Experimental procedure*

Our paradigm used an old/new design. During the training phase, participants were asked to memorize multiple successively presented images displayed for 6 s each. In each session, images were randomly selected, without replacement, from the stimulus database. To ensure that participants attended to the images, we also asked participants to rate the perceived beauty of each image on a scale from 1 to 5. During the test phase, the set of training images was augmented with an equal number of novel distractor images. Participants had to indicate whether they had previously seen each of the images in this test set. Participants' verbal responses were recorded by the experimenter. The test session was conducted 5 min after the conclusion of the memorization session. Images were shown on a 17-in. display under program control using MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard and Vision, 1997). The viewing distance was 40 cm, and stimulus images subtended 15° of visual angle horizontally and vertically.

To assess memory capacity, we progressively increased the number of images a participant was asked to memorize. The cardinalities we used were 10, 20, 40, and 80. At each cardinality, participants went through the memorization and test sessions before proceeding to the next higher cardinality. During the last postoperative check of newly sighted individuals and during recordings on blur-matched controls, each participant was exposed to eight different experimental conditions: four cardinalities (10, 20, 40, and 80) × two image types (real, abstract). Because only a few of the newly sighted patients carried out the experiment on set sizes of 80 at preoperative and early postoperative time points (the high-cardinality task was found to be exceedingly difficult at these time points), longitudinal within-subject analyses were restricted to a combination of three cardinalities (10, 20, and 40) and the two image types. The full combination of four set cardinalities and the two image types were taken into account for a comparison of newly sighted patients 1 year after surgery and control participants. Training and distractor images were unique across all experimental conditions (no images were repeated across cardinalities). No feedback was provided during the testing. The experiment lasted 90 min on average, and participants were free to take breaks between blocks.

The performance of each participant was characterized by calculating d′ using the number of hits, misses, correct rejections, and false alarms. Longitudinal-assessment data were collected at six different time points: Once before surgery, once following each eye surgery (which were typically 1 week apart), and 1 month, 6 months, and 12 months following surgery. Not all children contributed data to all longitudinal time points; challenges of travel from their rural domiciles to our center sometimes prevented them from participating in follow-up sessions. When a child was unable to perform the experimental task, as evidenced by unvarying responses to more than 10 stimuli in a sequence, the child's performance for that session was marked as equivalent to chance. Further, when a child was not able to complete all experimental conditions for a given time point, the data were disregarded for our analyses. Control participants were tested while they wore blur goggles simulating Snellen visual acuities of 20/200 and 20/500 (nonoverlapping groups of five and seven children at the two blur levels, respectively). These two acuity values, 20/200 and 20/500, approximated the range of postoperative acuities of the newly sighted patients—nine of the 12 patients were strictly within this acuity range, and three had marginally lower acuities (20/511, 20/524, and 20/543). Relative to 20/500, these three values correspond to differences in the logarithm of the minimum angle of resolution (logMAR) of less than 0.04, or less than one line of acuity measure, which can be considered

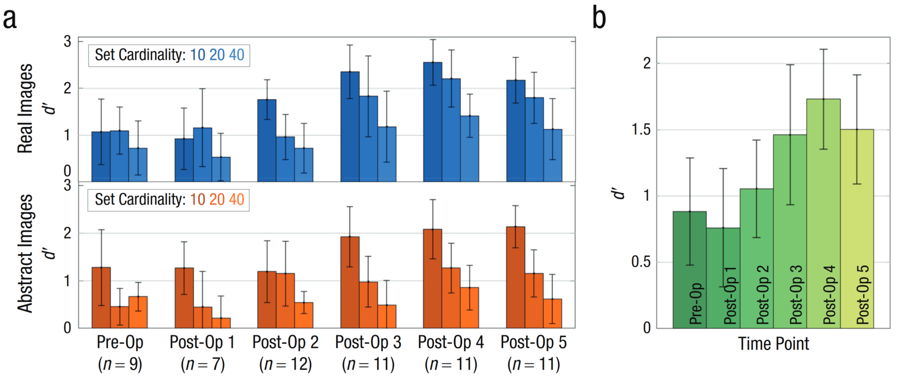


Figure 6.2: Recall performance. Averaged sensitivity (d′) is shown in (a) for real (top) and abstract (bottom) images as a function of set cardinality (10, 20, 40) and time point relative to surgery for congenital blindness. Participants were tested once before surgery (pre-op), once following surgery on each eye (Post-Op 1 and Post-Op 2, which were typically 1 week apart), and 1 month (Post-Op 3), 6 months (Post-Op 4), and 12 months (Post-Op 5) following surgery. Recall performance is shown in (b) as a function of time point, averaged across set cardinalities and image types. Error bars in both panels represent 95% confidence intervals.

clinically insignificant (Jones et al., 2003; Smith, 2006). Blurring was achieved by attaching Bangerter occlusion foils (Odell et al., 2008) to clear safety goggles. Assessing the performance of normally sighted participants while they were wearing blurring goggles enabled us to titrate the effects of reduced acuity to be comparable with the newly sighted children's postoperative outcomes, separate from nonoptical factors on visual memory performance.

6.6 RESULTS

As has been observed in previous studies of individuals with late sight onset (Ganesh et al., 2014; Kalia et al., 2017), visual outcomes across members of the experimental group can be quite variable. Despite this variability, however, some general trends are apparent, allowing for statistically meaningful inferences. Figure 6.2 summarizes the newly sighted individuals' data across time points (preoperative and multiple postoperative assessments), set cardinalities (10, 20, and 40), and stimulus types (real and abstract images).

To assess whether the visual working memory capacity of newly sighted children would longitudinally improve following surgery and whether such improvement may be due to the semantic content of an image, we examined within-subject performance variability as a function of set cardinality, image type, and postoperative recording time (see Table 6.1). First, we found that performance was significantly higher for real, compared with abstract, images ($p < .001$) as well as for lower, compared with higher, set cardinalities ($p < .001$). Crucially, the analyses revealed a positive relationship between the number of weeks following surgery and overall task performance ($p < .001$), quantifying the learning process visualized in Figure 6.2. The relationship between postsurgical recording time and performance further showed a significant negative interaction ($p = .017$) with the set cardinality (i.e., the temporal improvement was more salient in lower set sizes) but no significant interaction ($p = .864$) with image type (i.e., not providing evidence for a direct link between learning and the image semantics).

Although the aforementioned tests indicate that for the newly sighted children, visual memory capacity improved after sight onset within the first post-op year, it is particularly noteworthy that postoperative performance soon after surgery shows no improvement relative to preoperative performance: Comparing performance before treatment with performance a week after treatment ("Pre-Op" and "Post-Op 1" in Fig. 6.2) showed no improvements (see 6.9 Notes). The onset of patterned vision, therefore, does not bring about an immediate enhancement of visual memory capacity. Instead, the visual memory capacity develops only over the ensuing months, during which children gain visual experience. To examine whether, despite early deprivation, these performance levels would eventually reach those of participants undergoing normal visual development, we compared the performance of newly sighted participants 1 year after surgery with that of control participants whose acuity was artificially reduced to match that of the newly sighted children.

Figure 6.3 depicts the data across groups (newly sighted patients 1 year after surgery, blur-matched control participants), set cardinalities (10, 20, 40, 80), and image type (real, abstract). Within- and between-subject analyses revealed similar patterns in both the newly sighted and control groups (see Fig. 6.3): Performance decreased as a function of cardinalities ($p < .001$; see Table 6.2) and was higher for real relative to abstract images ($p < .001$). However, no significant difference emerged between the two groups ($p = .609$), and no significant interaction effects between the groups and set cardinality ($p = .928$) or image type ($p = .085$) were found. Thus, although there appear to be differences for specific combinations of image types and set cardinalities (e.g., real images with a set cardinality of 80), these differences did not reach statistical significance.

Thus, although the previous analysis (see Fig. 6.2 and Table 6.1) established that sight onset in itself is not sufficient to induce immediate visual memory improvements, here we found that visual experience for 1 year leads to visual memory patterns that are approximately comparable with those of control participants.

We also examined possible links between improvements in memory performance and changes in visual acuity. As documented in previous

Table 6.1: Results of Linear Mixed Models Examining Within-Subject Performance Variations of Newly Sighted Patients With Regard to Image Type (Real vs. Abstract), Time Point of Recording, and Set Cardinality (10, 20, 40) Note: The table shows unstandardized estimates of fixed effects. Significant results are indicated in boldface. CI = confidence interval.

Parameter	Estimate	SE	95% CI	df	t	p
Image type	**0.343**	**0.084**	**[0.178, 0.509]**	**348.033**	**4.082**	**< .001**
Time point	**0.020**	**0.005**	**[0.011, 0.029]**	**348.082**	**4.234**	**< .001**
Cardinality	**-0.027**	**0.003**	**[-0.034, -0.020]**	**348.033**	**-8.013**	**< .001**
Time Point × Image Type	-0.001	0.006	[-0.013, 0.011]	348.033	-0.172	.864
Time Point × Cardinality	**-0.0004**	**0.0002**	**[-0.0008, -0.0001]**	**348.033**	**-2.400**	**.017**
Time Point × Image Type × Cardinality	0.0002	0.0002	[-0.0002, 0.0006]	348.033	0.926	.355

Table 6.2: Results of Linear Mixed Models Examining Group-Relevant Performance Variations With Regard to Image Type (Within Subjects), Set Cardinality (Within Subjects), and Group (Between Subjects) Note: The table shows unstandardized estimates of fixed effects. Significant results are indicated in boldface. CI = confidence interval.

Parameter	Estimate	SE	95% CI	df	t	p
Image type	**0.905**	**0.206**	**[0.498, 1.313]**	**155**	**4.389**	**< .001**
Group	0.152	0.295	[-0.439, 0.742]	58.970	0.514	.609
Cardinality	**-0.019**	**0.003**	**[-0.026, -0.013]**	**155**	**-6.103**	**< .001**
Group × Image Type	-0.517	0.298	[-1.106, 0.072]	155	-1.734	.085
Group × Cardinality	-0.0004	0.005	[-0.010, 0.009]	155	-0.091	.928
Group (Experimental) × Image Type × Cardinality	-0.001	0.005	[-0.010, 0.008]	155	-0.228	.820
Group (Control) × Image Type × Cardinality	-0.0002	0.004	[-0.009, 0.009]	155	-0.037	.971

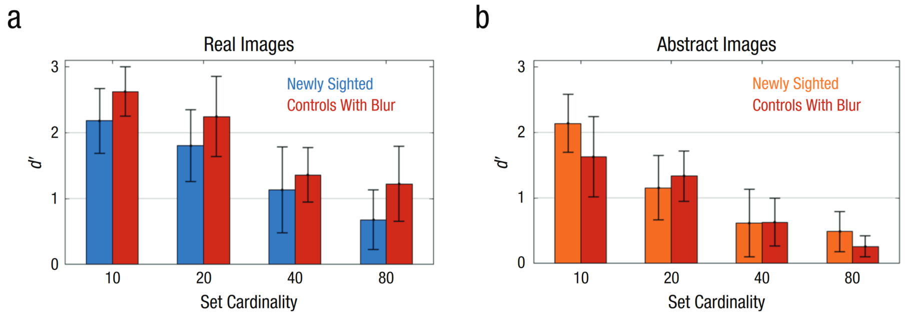


Figure 6.3: Recall performance on the final memory task (12 months after surgery). Averaged sensitivity (d′) is shown for (a) real and (b) abstract images as a function of set cardinality (10, 20, 40, 80), separately for newly sighted children and blur-matched control participants. Error bars represent 95% confidence intervals.

work, visual acuity shows modest improvements over time following sight-restoring surgery (Ganesh et al., 2014). The same was true of this cohort of children. Figure 6.7 in the Supplemental Material shows the change in acuity and d′ as a function of time after surgery for the 12 newly sighted patients. The dashed red line depicts performance of control participants at a blur level of 20/500. The point to note here is that although patients' acuity and d′ showed an improvement over time, it took several months for their performance (d′) to be at par with that of control participants, despite the fact that the induced blur level in the latter was worse than the postoperative acuity of most of the patients. Thus, there is a protracted temporal progression over which visual memory performance develops to a level achievable by control participants with comparable or worse induced acuity. For higher cardinalities, performance of many of the patients remained below that of control participants even though their acuity was better than 20/500. Hence, although the newly sighted Prakash participants did experience longitudinal changes in their postoperative acuity, these changes did not provide an adequate account of the improvement in their visual memory skills. Specifically, memory performance of the Prakash children early in the postsurgical timeline was lower than what was feasible in principle given their acuity.

Figure 6.5 in the Supplemental Material shows a color-coded matrix of within-subject correlations of acuity and performance over time for all participants in the experimental group. Figure 6.6 shows the correlations of acuity and recognition performance at the final time point measured for all participants.

6.7 DISCUSSION

The current study explored the impact of early visual deprivation on the later development of visual memory recall once sight is restored.

The primary finding was that the basic ability to recall previously seen images is initially poor at sight onset but improves significantly in the ensuing months. Despite initially poor post-treatment performance on all image types and cardinalities, the improvements that emerged over time ultimately showed superior recall of real compared with abstract images and an overall pattern that was graded as a function of increased cardinality. Following a year of visual experience, patients' overall performance still fell slightly, though not significantly, short of the typically developing control participants' performance, so we are not ruling out the possibility that early visual deprivation may limit the development of visual memory later in life. However, it is notable that the effects of stimulus type and cardinality were found for both groups of participants, suggesting that some aspects of the functional organization of visual memory that emerge following early visual deprivation may be similar to those that emerge from typical visual development. Importantly, we found that the formation of high-fidelity memory representations for semantically meaningful information emerges slowly and depends on visual experience—a result that is consistent with recent models of visual memory representations in which information about real-world scenes is stored as a hierarchical feature bundle with object-level information at the top and basic features at the bottom (Brady et al., 2011).

The significant longitudinal improvement found in memory capacity suggests that this ability is at least partly resilient to protracted periods of visual deprivation. What might account for this resilience? It is believed that rather than being a localized process, memory is subserved by multiple cortical areas. Candidate areas include the prefrontal cortex (Runyan et al., 2004) and the lateral occipital complex (Gayet et al., 2018), among others. Given this distributed notion of memory as drawing on a distributed network, one possible explanation of the observed resilience is that substrates of visual memory that are located in higher cortical areas may be less susceptible to deprivation than early sensory cortices, such as V1. A related possibility is that at least some memory resources may be amodal, accessible to multiple sensory modalities and cognitive processes (Loomis et al., 2012; McCarthy and Warrington, 1988; Schumacher et al., 1996). Such amodal stores can be maintained by modalities other than vision while an individual is blind and thus lessen their vulnerability to absence of visual information. Additionally, it is worth noting that the newly sighted patients were not entirely denied visual stimulation prior to surgery. Although profoundly visually compromised, they did experience light stimulation and even rudimentary pattern perception. Such experience, however limited, may play a role in sustaining the neural substrates for visual memory. The question of whether total removal of visual stimulation for an extended period after birth can still be followed by the development of visual memory can potentially be investigated with nonhuman animal studies using dark-rearing regimens that have been extensively employed in the past (e.g., Fagiolini et al., 1994).

Beyond the evidence of resilience, a more specific question concerns the nature of mechanisms that might account for the observed enhancement in memory performance. There are at least two broad possibilities. The first involves increases in intrinsic memory capacity, perhaps through the recruitment of greater neural resources for information storage. The second relies on representational elaboration; increasing richness of the representational vocabulary for visual content would render images more discriminable and would become manifest as improved memory performance even without changes in the storage capacity per se. The rapid rate of improvement we observed in the newly sighted children and the relatively advanced age at which such improvements were happening lead us to favor the latter account over the former. We have conducted computational simulations in order to verify the plausibility of the idea that increasing experience with patterned imagery would lead to progressive elaborations of internal representations of individual images, rendering them more discriminable and thereby achieving steadily better performance on the kind of memory task that we have used with the newly sighted children.

To this end, we trained a prototypical convolutional neural network, the AlexNet (Krizhevsky et al., 2012), on the Caltech-101 object database (Fei-Fei et al., 2007), with individual images rescaled and cropped to 100 × 100 pixels. We then analyzed its internal representations throughout training. Specifically, after 1, 5, 10, 20, and 50 epochs of training, we presented 100 new exemplar images to the (partially) trained network and recorded the resulting activations at five different layers of the network (see Fig. 6.4a). We then applied a multidimensional scaling (MDS) analysis to visualize the activations for each image in a two-dimensional space, separately for each of the five layers and each of the five different numbers of epochs. With this analysis, we investigated representational elaboration: whether, and where in the network, continued training yielded more diverse visual representations that consequently rendered images more discriminable.

Figure 6.4b shows the results. A modest amount of representational elaboration can be seen in the later convolutional layers, although no such elaboration is evident in the first layer. The elaboration is particularly notable in the fully connected layers; more training renders images more discriminable.

While in the first fully connected layer, continued elaboration can be seen even in comparatively late epochs (e.g., between Epochs 20 and 50), in the second fully connected layer, the representation appears more elaborate from early on. With the softmax operation applied to

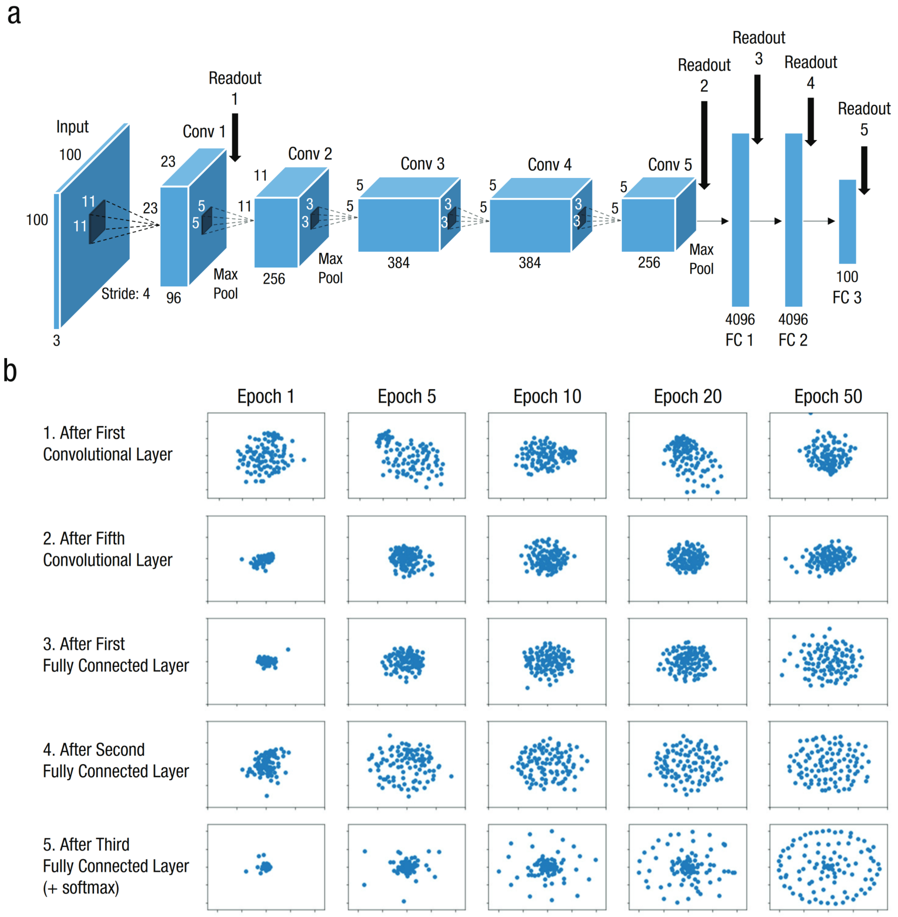


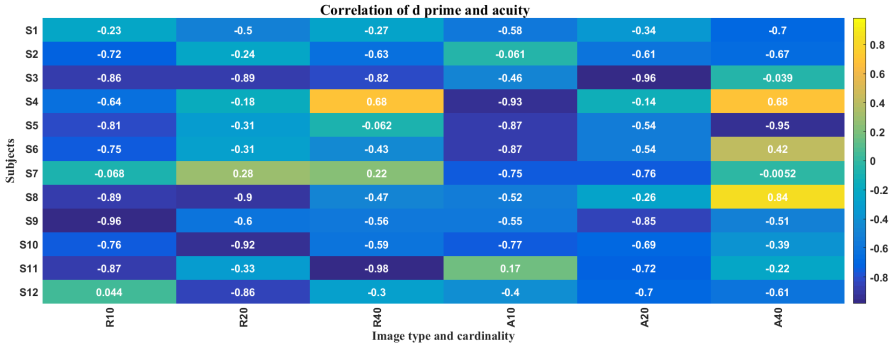
Figure 6.4: The convolutional network, the AlexNet, used to conduct computational simulations. The schematic (a) illustrates the functioning of the network. Arrows indicate the five readout sites. "Conv" refers to convolutional layers; "FC" refers to fully connected layers. Multidimensional scaling–derived visualizations of the activities of units in the indicated readout sites are shown in (b). With the exception of the last readout site, to which the softmax operation had already been applied, the activity magnitudes have been normalized. A wider spread in the 2D plots depicting the projected space can therefore not be the result of larger absolute values that may have simply been reinforced during training. The axis scales are identical across epochs but not across layers.

the last fully connected layer, we see that task-relevant fine-tuning keeps proceeding even until later epochs.

The results from these computational simulations lend credence to the possibility that the progressive improvements in the performance of newly sighted children on the old/new task may have arisen from the increased representational elaborations of images as the children steadily gained more visual experience.

We need to keep in mind some important caveats when interpreting the results of this study. First, although the experimental data clearly show improvements in visual memory performance following sight onset, the mechanisms underlying this improvement remain unclear. These may include intrinsic changes in the memory subsystems of the brain and/or changes in representational richness of the early visual areas (as suggested by the computational simulations). Related to the second possibility is the potential contribution of acuity improvements. Given the changes in acuity over the same timeline as that of improvements in visual memory performance, we cannot rule out an influence of the former on the latter; increased acuity would permit better discrimination between images by providing access to a richer set of spatial features. Although acuity changes may well contribute to progression in visual memory performance, it is unclear whether they constitute a complete account of the observed improvements. As mentioned in the Results section, newly sighted people exhibit residual deficits in memory performance relative to control participants with induced acuity loss. Control participants' resilience to acuity reduction could arise from the more extensive visual experience they have had with normal resolution imagery, enabling them to better interpret degraded inputs. Not having had the benefit of such experience, the newly sighted group's performance may be more acuity limited than the control group's. Second, accumulating postoperative experience with the real world leads to an increase in categorization abilities (Ostrovsky et al., 2009). The instantiation of visual categories may play a role in organizing visual inputs and thus may impact their memorizability. Third, the patient population we worked with included adolescents and young adults. It is unclear whether these results can be extrapolated to individuals who gain sight much later in life, as was the case with the individuals described in earlier reports (Fine et al., 2003; Gregory and Wallace, 1963; Sacks, 1995; Valvo et al., 1971).

In summary, the progressive improvement in memory performance after sight onset leads us to conclude that the substrates responsible for visual memory retain at least some measure of plasticity despite several years of visual deprivation. It is unclear, though, whether these improvements result from increases in storage per se or from enhancements in image encoding, perhaps driven by improvements in visual acuity. Furthermore, additional studies are needed to de-



	R10	R20	R40	A10	A20	A40
S1	-0.23	-0.5	-0.27	-0.58	-0.34	-0.7
S2	-0.72	-0.24	-0.63	-0.061	-0.61	-0.67
S3	-0.86	-0.89	-0.82	-0.46	-0.96	-0.039
S4	-0.64	-0.18	0.68	-0.93	-0.14	0.68
S5	-0.81	-0.31	-0.062	-0.87	-0.54	-0.95
S6	-0.75	-0.31	-0.43	-0.87	-0.54	0.42
S7	-0.068	0.28	0.22	-0.75	-0.76	-0.0052
S8	-0.89	-0.9	-0.47	-0.52	-0.26	0.84
S9	-0.96	-0.6	-0.56	-0.55	-0.85	-0.51
S10	-0.76	-0.92	-0.59	-0.77	-0.69	-0.39
S11	-0.87	-0.33	-0.98	0.17	-0.72	-0.22
S12	0.044	-0.86	-0.3	-0.4	-0.7	-0.61

Figure 6.5: (Supplemental Figure) A matrix showing correlations for each of the 12 experimental group participants between their longitudinally assessed visual acuity and memory performance on each of the six experimental conditions (image-type, set cardinality).

termine whether the extent of available plasticity is modulated by age at treatment, whether a correlation exists between nonvisual and visual memory capacity, and what neural changes accompany the development of visual memory in the newly sighted.

6.8 supplemental material

6.8.1 *Participant information*

Experimental group: 12 children with dense congenital bilateral cataracts participated in this study. All were drawn from rural areas in north India. Although their parents noticed their visual impairments within the first six months after birth, none of the children received treatment because medical facilities were not available locally and the families could not afford care in city hospitals. They were identified as candidates for treatment during outreach sessions for pediatric ophthalmic screening. The field-based screening was followed by a thorough ophthalmic examination at our medical center in New Delhi using direct and indirect ophthalmoscopes, slit lamps and B-scan ultrasonography. This examination assessed ocular pathologies in anterior and posterior eye segments and was undertaken in conjunction with standard tests of visual function. For the cases described here, the pathology was confined to the lens bilaterally. Pre-operative vision was limited to the perception of hand-movements close to the face or very limited pattern vision. All children underwent cataract removal surgery and an intraocular lens (IOL) implant. Post-treatment, the children achieved resolution-acuities for near viewing ranging from 20/138 to 20/543 Snellen (please see Table 6.3 for all acuity values). Acuities were assessed via the use of Landolt C patterns.

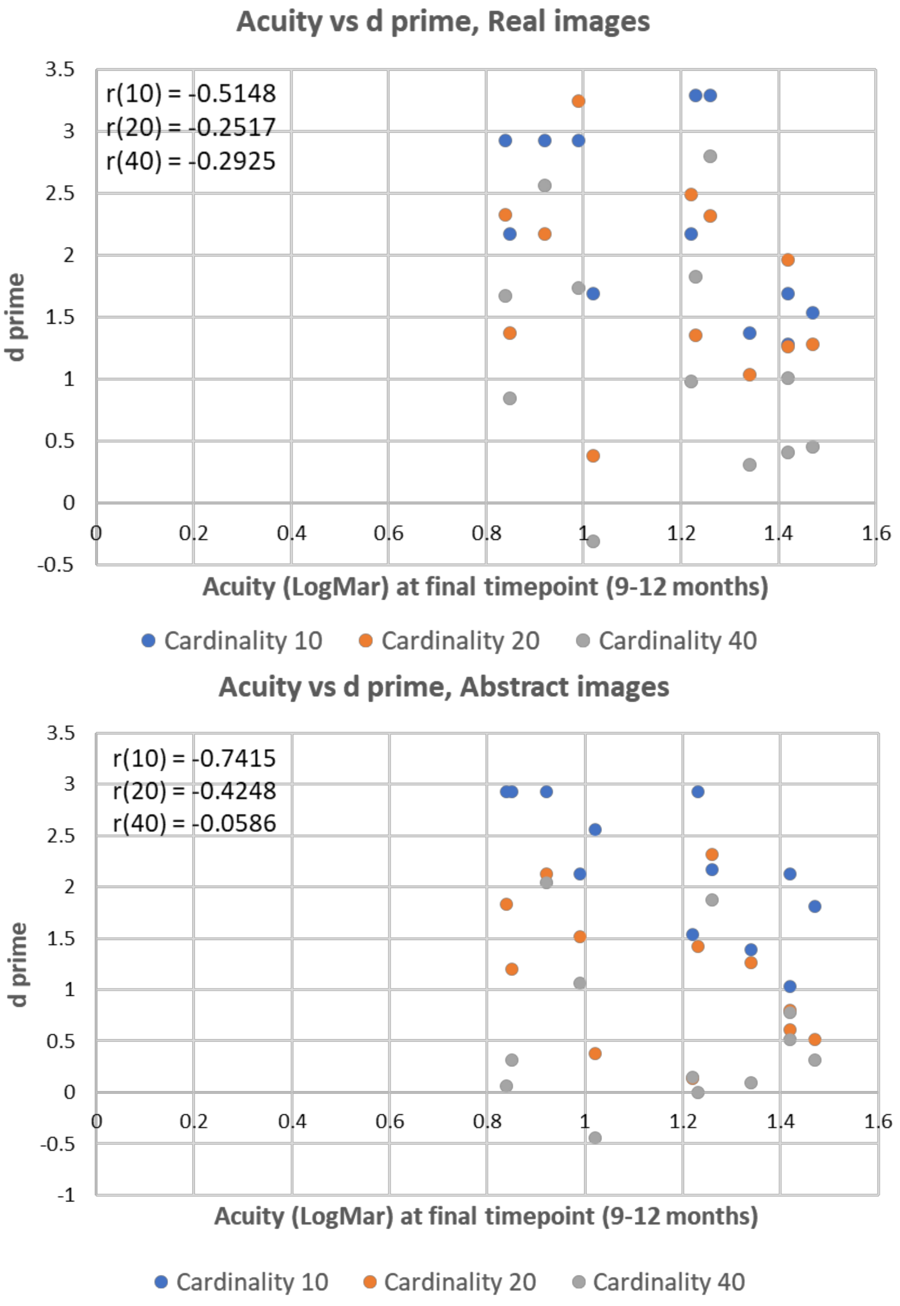


Figure 6.6: (Supplemental Figure) Scatterplots showing d-prime versus final recorded acuity across all experimental group participants in each of the six assessment conditions. The r values are shown in the plots. The corresponding p values are as follows: for real images $p(10) = 0.0868$, $p(20) = 0.4301$, $p(40) = 0.3562$ and for abstract images $p(10) = 0.0058$, $p(20) = 0.1686$, $p(40) = 0.8564$.

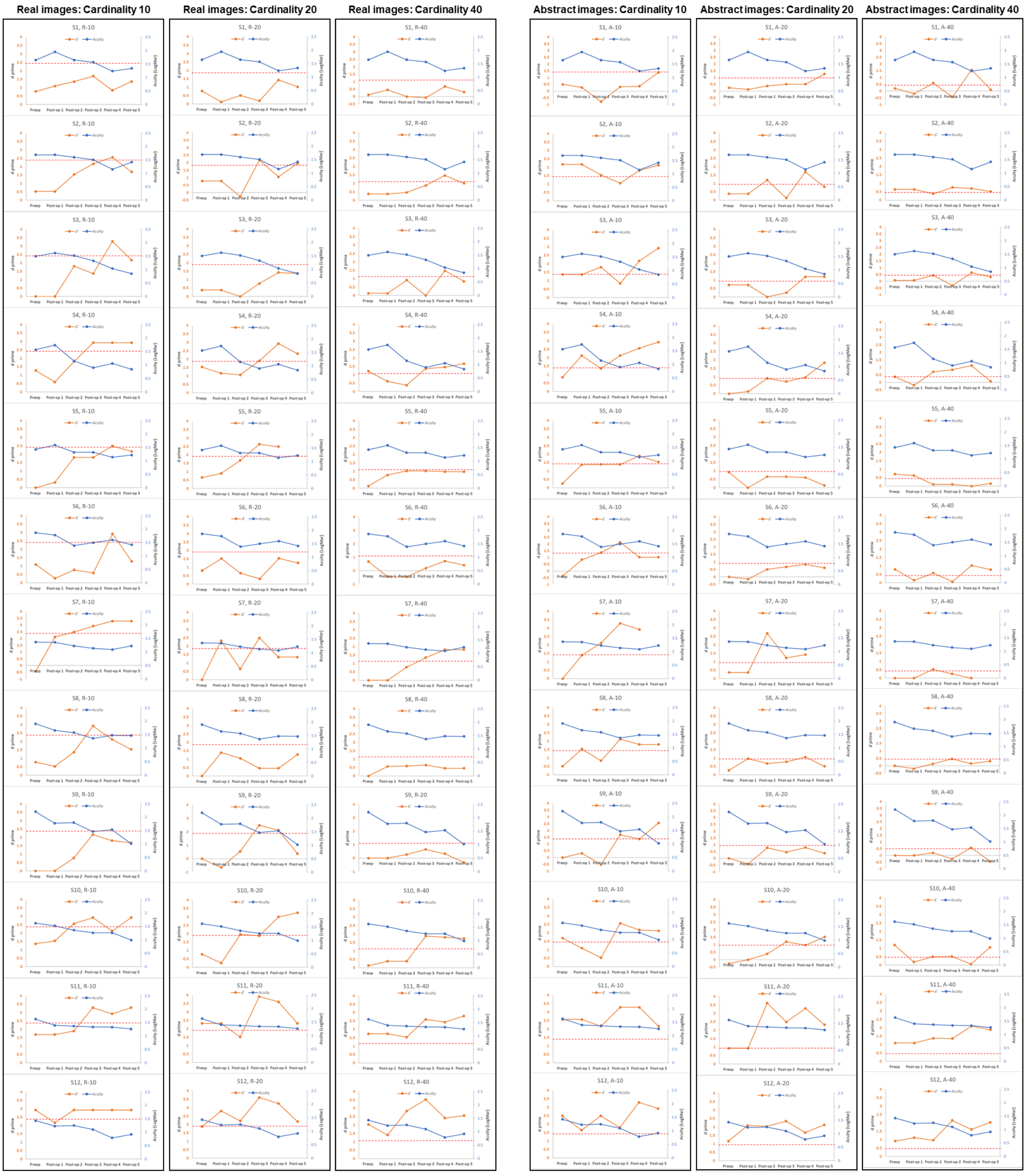


Figure 6.7: (Supplemental Figure) For better resolution, please refer to https://doi.org/10.25384/SAGE.19739211.v1

Table 6.3: (Supplemental Table) Control group: 12 children, recruited from a school in New Delhi served as controls. They had normal or corrected to normal vision and similar socio-economic backgrounds as the newly-sighted children. Informed consent/assent was obtained from all children and their guardians.

Subjects	Age	Gender	Pre-operative acuity	Post-operative acuity (Snellen)
S1	22	M	20/889	20/447
S2	15	M	20/969	20/524
S3	11	F	20/627	20/143
S4	13	M	20/751	20/138
S5	9	F	20/541	20/335
S6	8	F	20/1477	20/543
S7	8	F	20/474	20/340
S8	17	M	20/1617	20/511
S9	12	M	20/3251	20/207
S10	12	M	20/839	20/197
S11	15	F	20/853	20/364
S12	11	F	20/535	20/165

ACKNOWLEDGEMENTS

We thank all the children and their parents who participated in this study, as well as all the members of our extended team, who were instrumental in identifying congenitally blind children with treatable conditions.

6.9 NOTES

It is worth noting that preoperative performance was not exactly at chance level. This is likely because even with dense cataracts, it is possible to get information about overall luminance and average color. This information may have been sufficient for getting some responses correct, especially when the cardinality was low, but this rudimentary characterization of the image was not very useful with larger cardinalities. Indeed, as is evident in Figures 6.2 and 6.3, with higher cardinality values, the preoperative d' values dropped close to zero (a similar decrease can be observed in control participants; see Fig. 6.3).

TRANSPARENCY

Action Editor: Angela Lukowski Editor: Patricia J. Bauer

Author Contributions

P. Gupta and P. Sinha developed the study concept. P. Gupta, S. Gilad-Gutnick, and P. Sinha designed the experimental protocol. P. Gupta, P. Shah, and K. Tiwari collected experimental data. P. Gupta, P. Shah, L. Vogelsang, and P. Sinha analyzed and interpreted the data. M. Vogelsang and L. Vogelsang performed the computational simulations. S. Ganesh conducted the surgical procedures and ophthalmic characterization of Prakash patients. P. Gupta, P. Shah, S. Gilad-Gutnick, M. Vogelsang, L. Vogelsang, T. Gandhi, and P. Sinha drafted the manuscript. All authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research was supported by the National Institute of Health (R01EY020517), the James McDonnell Foundation, the Rotary Club of New Delhi, the Nick Simons Foundation, the Halis Family Foundation, and the Sikand Foundation.

Open Practices

Data and materials for this study have not been made publicly available, and the design and analysis plans were not preregistered.

REFERENCES

American Foundation for the Blind (2020). *Low vision and legal blindness terms and descriptions*. https://www.afb.org/blindness-and-low-vision/eye-conditions/low-vision-and-legal-blindness-terms-and-descriptions.

Bach, Michael et al. (1996). “The Freiburg Visual Acuity Test-automatic measurement of visual acuity.” In: *Optometry and vision science* 73.1, pp. 49–53.

Bellhouse-King, Mathew W and Lionel G Standing (2007). "Recognition memory for concrete, regular abstract, and diverse abstract pictures." In: *Perceptual and motor skills* 104.3, pp. 758–762.

Braddick, Oliver, Janette Atkinson, and John Wattam-Bell (2003). "Normal and anomalous development of visual motion processing: motion coherence and 'dorsal-stream vulnerability'." In: *Neuropsychologia* 41.13, pp. 1769–1784.

Brady, Timothy F, Talia Konkle, and George A Alvarez (2011). "A review of visual memory capacity: Beyond individual items and toward structured representations." In: *Journal of vision* 11.5, pp. 4–4.

Brady, Timothy F, Talia Konkle, George A Alvarez, and Aude Oliva (2008). "Visual long-term memory has a massive storage capacity for object details." In: *Proceedings of the National Academy of Sciences* 105.38, pp. 14325–14329.

Brainard, David H and Spatial Vision (1997). "The psychophysics toolbox." In: *Spatial vision* 10.4, pp. 433–436.

Cobb, SR and CF MacDonald (1978). "Resolution acuity in astigmats: evidence for a critical period in the human visual system." In: *The British journal of physiological optics* 32, pp. 38–49.

Cowan, Nelson, Angela M AuBuchon, Amanda L Gilchrist, Timothy J Ricker, and J Scott Saults (2011). "Age differences in visual working memory capacity: Not based on encoding limitations." In: *Developmental science* 14.5, pp. 1066–1074.

Ellemberg, Dave, Terri L Lewis, Daphne Maurer, Sonia Brar, and Henry P Brent (2002). "Better perception of global motion after monocular than after binocular deprivation." In: *Vision research* 42.2, pp. 169–179.

Fagiolini, Michela, Tommaso Pizzorusso, Nicoletta Berardi, Luciano Domenici, and Lamberto Maffei (1994). "Functional postnatal development of the rat primary visual cortex and the role of visual experience: dark rearing and monocular deprivation." In: *Vision research* 34.6, pp. 709–720.

Fei-Fei, Li, Rob Fergus, and Pietro Perona (2007). "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories." In: *Computer Vision and Image Understanding* 106.1. Special issue on Generative Model Based Vision, pp. 59–70. ISSN: 1077-3142.

Ferrara, Katrina, Sarah Furlong, Soojin Park, and Barbara Landau (2017). "Detailed visual memory capacity is present early in childhood." In: *Open Mind* 2.1, pp. 14–25.

Fine, Ione, Alex R Wade, Alyssa A Brewer, Michael G May, Daniel F Goodman, Geoffrey M Boynton, Brian A Wandell, and Donald IA MacLeod (2003). "Long-term deprivation affects visual perception and cortex." In: *Nature neuroscience* 6.9, pp. 915–916.

Ganesh, Suma, Priyanka Arora, Sumita Sethi, Tapan K Gandhi, Amy Kalia, Garga Chatterjee, and Pawan Sinha (2014). "Results of late surgical intervention in children with early-onset bilateral cataracts." In: *British Journal of Ophthalmology* 98.10, pp. 1424–1428.

Gayet, Surya, Chris LE Paffen, and Stefan Van der Stigchel (2018). "Visual working memory storage recruits sensory processing areas." In: *Trends in cognitive sciences* 22.3, pp. 189–190.

Gregory, Richard L and Jean G Wallace (1963). "Recovery from early blindness." In: *Experimental psychology society monograph* 2, pp. 65–129.

Hubel, David H and Torsten N Wiesel (1970). "The period of susceptibility to the physiological effects of unilateral eye closure in kittens." In: *The Journal of physiology* 206.2, pp. 419–436.

Jones, Deborah, Carol Westall, Karin Averbeck, and Mohamed Abdolell (2003). "Visual acuity assessment: a comparison of two tests for measuring children's vision." In: *Ophthalmic and Physiological Optics* 23.6, pp. 541–546.

Kalia, Amy, Tapan Gandhi, Garga Chatterjee, Piyush Swami, Harvendra Dhillon, Shakeela Bi, Naval Chauhan, Shantanu Das Gupta, Preeti Sharma, Saahil Sood, et al. (2017). "Assessing the impact of a program for late surgical intervention in early-blind children." In: *Public Health* 146, pp. 15–23.

Kalia, Amy, Luis Andres Lesmes, Michael Dorr, Tapan Gandhi, Garga Chatterjee, Suma Ganesh, Peter J Bex, and Pawan Sinha (2014). "Development of pattern vision following early and extended blindness." In: *Proceedings of the National Academy of Sciences* 111.5, pp. 2035–2039.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems* 25.

Le Grand, Richard, Catherine J Mondloch, Daphne Maurer, and Henry P Brent (2001). "Early visual experience and face processing." In: *Nature* 410.6831, pp. 890–890.

Lewis, Terri L and Daphne Maurer (2005). "Multiple sensitive periods in human visual development: evidence from visually deprived children." In: *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 46.3, pp. 163–183.

Loomis, Jack M, Roberta L Klatzky, Brendan McHugh, and Nicholas A Giudice (2012). "Spatial working memory for locations specified by vision and audition: Testing the amodality hypothesis." In: *Attention, Perception, & Psychophysics* 74, pp. 1260–1267.

Madigan, Stephen (2014). "Picture memory." In: *Imagery, memory and cognition*, pp. 65–89.

Maurer, Daphne, Terri L Lewis, and Catherine J Mondloch (2005). "Missing sights: consequences for visual cognitive development." In: *Trends in cognitive sciences* 9.3, pp. 144–151.

Maurer, Daphne, Catherine J Mondloch, and Terri L Lewis (2007). "Effects of early visual deprivation on perceptual and cognitive development." In: *Progress in brain research* 164, pp. 87–104.

McCarthy, Rosaleen A and Elizabeth K Warrington (1988). "Evidence for modality-specific meaning systems in the brain." In: *Nature* 334.6181, pp. 428–430.

Nickerson, Raymond S (1965). "Short-term memory for complex meaningful visual configurations: A demonstration of capacity." In: *Canadian Journal of Psychology/Revue canadienne de psychologie* 19.2, p. 155.

Odell, Naomi V, David A Leske, Sarah R Hatt, Wendy E Adams, and Jonathan M Holmes (2008). "The effect of Bangerter filters on optotype acuity, Vernier acuity, and contrast sensitivity." In: *Journal of American Association for Pediatric Ophthalmology and Strabismus* 12.6, pp. 555–559.

Ostrovsky, Yuri, Aaron Andalman, and Pawan Sinha (2006). "Vision following extended congenital blindness." In: *Psychological Science* 17.12, pp. 1009–1014.

Ostrovsky, Yuri, Ethan Meyers, Suma Ganesh, Umang Mathur, and Pawan Sinha (2009). "Visual parsing after recovery from blindness." In: *Psychological Science* 20.12, pp. 1484–1491.

Riggs, Kevin J, James McTaggart, Andrew Simpson, and Richard PJ Freeman (2006). "Changes in the capacity of visual working memory in 5-to 10-year-olds." In: *Journal of experimental child psychology* 95.1, pp. 18–26.

Rose, Susan A, Judith F Feldman, and Jeffery J Jankowski (2001). "Visual short-term memory in the first year of life: capacity and recency effects." In: *Developmental psychology* 37.4, p. 539.

Ross-sheehy, Shannon, Lisa M Oakes, and Steven J Luck (2003). "The development of visual short-term memory capacity in infants." In: *Child development* 74.6, pp. 1807–1822.

Runyan, Jason D, Anthony N Moore, and Pramod K Dash (2004). "A role for prefrontal cortex in memory storage for trace fear conditioning." In: *Journal of Neuroscience* 24.6, pp. 1288–1295.

Sacks, Oliver (1995). *An anthropologist on Mars: Seven paradoxical tales*. Vintage Books.

Schumacher, Eric H, Erick Lauber, Edward Awh, John Jonides, Edward E Smith, and Robert A Koeppe (1996). "PET evidence for an amodal verbal working memory system." In: *Neuroimage* 3.2, pp. 79–88.

Shepard, Roger N (1967). "Recognition memory for words, sentences, and pictures." In: *Journal of verbal Learning and verbal Behavior* 6.1, pp. 156–163.

Simmering, Vanessa R (2012). "The development of visual working memory capacity during early childhood." In: *Journal of experimental child psychology* 111.4, pp. 695–707.

Sinha, Pawan (2013). "Once blind and now they see." In: *Scientific American* 309.1, pp. 48–55.

Sinha, Pawan and Richard Held (2012). “Sight restoration.” In: *F1000 Medicine Reports* 4.

Slotnick, Scott D (2004). “Visual memory and visual perception recruit common neural substrates.” In: *Behavioral and Cognitive Neuroscience Reviews* 3.4, pp. 207–221.

Smith, George (2006). “Refraction and visual acuity measurements: what are their measurement uncertainties?” In: *Clinical and Experimental Optometry* 89.2, pp. 66–72.

Standing, Lionel (1973). “Learning 10000 pictures.” In: *Quarterly Journal of Experimental Psychology* 25.2, pp. 207–222.

Standing, Lionel, Jerry Conezio, and Ralph Norman Haber (1970). “Perception and memory for pictures: Single-trial learning of 2500 visual stimuli.” In: *Psychonomic science* 19.2, pp. 73–74.

Tusa, Ronald J, MX Repka, Carolyn B Smith, and SJ Herdman (1991). “Early visual deprivation results in persistent strabismus and nystagmus in monkeys.” In: *Investigative ophthalmology & visual science* 32.1, pp. 134–141.

Valvo, Alberto, Leskie L Clark, and Zofja Z Jastrzembska (1971). “Sight restoration after long-term blindness: The problems and behavior patterns of visual rehabilitation.” In: *(American Foundation for the Blind)*.

Von Senden, Marius (1960). *Space and sight: the perception of space and shape in the congenitally blind before and after operation.* Free Press of Glencoe.

Walker, Peter, Graham J Hitch, Alison Doyle, and Tracey Porter (1994). “The development of short-term visual memory in young children.” In: *International Journal of Behavioral Development* 17.1, pp. 73–89.

Wolbers, Thomas, Roberta L Klatzky, Jack M Loomis, Magdalena G Wutte, and Nicholas A Giudice (2011). “Modality-independent coding of spatial layout in the human brain.” In: *Current Biology* 21.11, pp. 984–989.

Yuille, John C (2014). *Imagery, memory and cognition (PLE: Memory): Essays in honor of Allan Paivio*. Psychology Press.

7 SCHOLASTIC STATUS OF CONGENITALLY BLIND CHILDREN FOLLOWING SIGHT SURGERY

Content from

Bi, S., Chawariya, A., Ganesh, S., Gupta, P., Huang, Y., Jazayeri, K., Kumar, R., Ralekar, C., Singh, C., Tiwary, A., Vogelsang, L., **Vogelsang, M.**, Yadav, M., & Sinha, P. (2023). "Scholastic status of congenitally blind children following sight surgery". **Published** in International Journal of Special Education, 37(2), 160-168. (alphabetical author ordering, except for P. Sinha)

7.1 ABSTRACT

India is home to a large population of blind children, many with treatable conditions. Project Prakash identifies treatable blind children and provides them with eye surgeries. Once treated, these children are given the opportunity to further their education. To understand their educational needs, we undertook a diagnostic screening exercise. Specifically, we designed math proficiency assessments and evaluated the scholastic preparation of 54 Prakash children across a broad age range. We found that the proficiency of most Prakash children was well below age-appropriate levels. Even those enrolled in high schools had assessed proficiencies around the 3rd-grade level. Furthermore, due to a lack of basic instruction on interpreting print material, many children continued using Braille even after gaining sight. The contrast that these findings present relative to the official standing of the children in their schools makes a compelling case for more rigorous assessments and better educational interventions for visually-impaired/sight-restored children. Also, we argue that these educational interventions should be coupled with visual function assessments to ensure that the presented material is accessible to the child. Our scholastic assessments, suitable for being administered over the phone, will be made available for use by other researchers and educationists.

7.2 KEYWORDS

Congenital blindness, special education, mathematics, diagnostic tests

7.3 BACKGROUND

We seek to understand the educational status of an unusual population of children. These are children who were born blind in India and, due

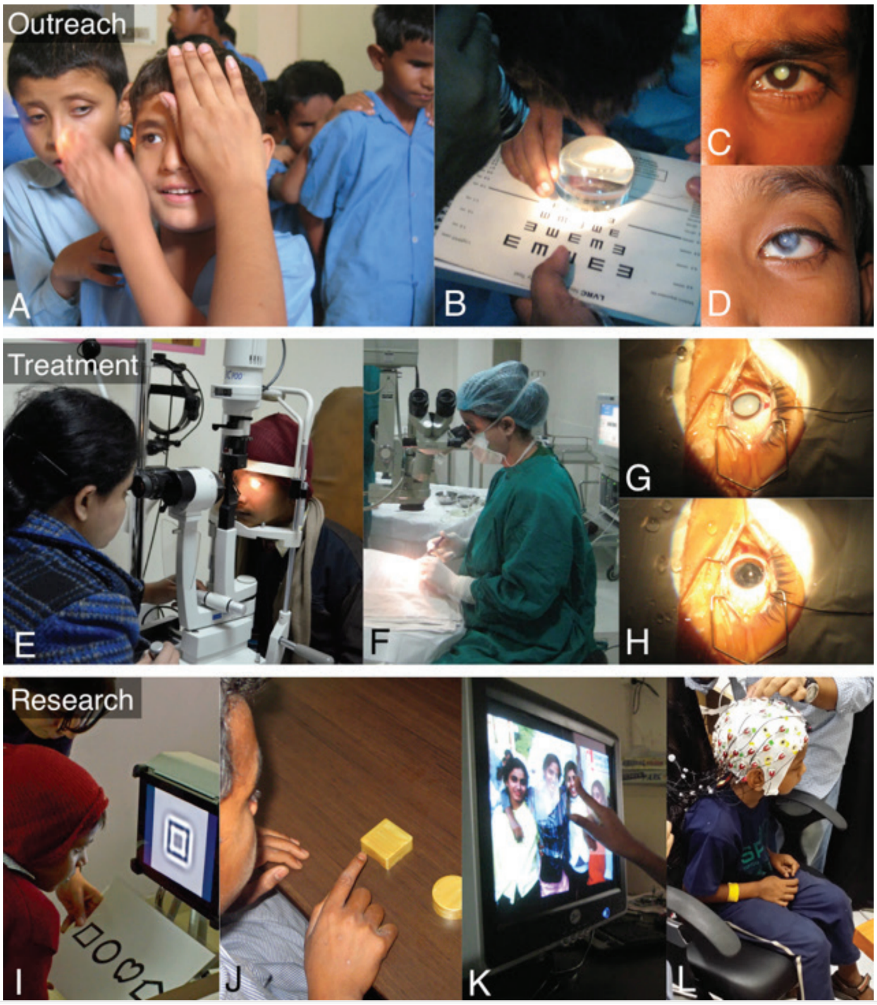


Figure 7.1: Project Prakash comprises three components: Outreach to identify children with treatable visual impairments and blindness (panels A-D); Treatment that typically involves pediatric cataract surgery (panels E-H); Scientific research to examine the development of vision following sight onset (panels I-L).

to financial or geographical reasons, stayed deprived of medical care, even though their blindness was treatable. Since 2005, an MIT-based initiative, Project Prakash, has actively searched for such children and provided them with eye surgeries (Sinha, 2013). In doing so, the project has attempted to address the pressing humanitarian need to treat blind children and has also gained scientifically valuable insights regarding how the brain adapts to the influx of visual information late in childhood. The work has positively impacted the quality of life of the treated children (Kalia et al., 2017) and has also uncovered several crucial results regarding the timelines and mechanisms of visual development (Gandhi et al., 2017; Gupta et al., 2022; Ostrovsky et al., 2009). Thus far, the project has brought sight to over 500 blind children. Figure 7.1 shows a few vignettes from Project Prakash.

We focus here on the educational trajectories of the treated children. In following up with the children, we found that even after gaining sight, several of them were unable to get admission into regular schools, which were unwilling to make accommodations for lingering visual deficits (such as sub-par visual acuity (Ganesh et al., 2014) and unfamiliarity with print). Hence, the Prakash children find it challenging to gain a foothold on the path that could lead them to eventual employment and financial independence. Deprived of opportunities for mainstream educational advancement, the children either continue at schools for the blind or, worse, are confined to home. Girls are often married off young, effectively extinguishing any possibility of their self-actualization.

Given the Prakash children's near-total lack of mainstream educational options, there is a clear need to provide them with a 'bridge' course. The figurative 'bridge' is between their current scholastic preparation (acquired through their enrollment in schools for the blind) and a middle-school level of proficiency. After progressing through such a course, children will have a better chance of being mainstreamed into conventional schools since they would not require remedial accommodations from these schools.

To design such a bridge course, a necessary first step is to assess the current state of scholastic preparation of the Prakash children. This is the goal we focus on in this report. Specifically, within mathematics, we describe the diagnostic tests we prepared for a range of grades, the procedures we followed in administering them, and, importantly, the results of these assessments. We have chosen to constrain our initial investigation to mathematics since the subject is essential for many real-world tasks (e.g., maintaining home accounts and shopping) and critical for success in other areas, such as science. The findings are sobering, revealing the dismally low level of scholastic readiness of the Prakash children despite having been enrolled in schools for the blind for multiple years. The results are instructive not just for our specific goal of determining what level to target our bridge course but, more broadly, about the need to examine how the current schooling system for the blind must be improved to make a substantive impact on the intellectual preparation of its students.

7.3.1 *Past work*

The newly sighted Prakash children constitute such a unique population that no prior work on their educational status exists in the literature. The reported data relate exclusively to untreated blind children. The findings, by and large, paint a disheartening picture. For most such children in India, education is a rare privilege. The paucity of schools catering to blind children is a significant bottleneck. Special schools for the blind can accommodate only about 8% of all blind children in the country (Rahi et al., 1995). The social stigma associated with the condition (Rohwerder, 2018) often induces parents to keep their blind children confined to home. As for the possibility of integrating severely visually impaired children in mainstream schools, the statistics are discouraging. According to a nationwide survey conducted in India by the state-run National Council for Educational Research and Training (NCERT), only 21.11% of all schools in India have provision for inclusive education for children with disabilities (Sreekanth, 2016). Even more sobering is the low rate of teacher training in special education. Only 1.32% of all teachers have any kind of training to work with children with disabilities. Given this lack of preparation of the schools and teachers, even for the children who gain access to a school, the actual acquisition of education is a challenge.

Mathematics education has been a particularly difficult challenge for the visually impaired (Bell and Silverman, 2019). The reliance on equations with special symbols not represented in Braille, as well as the reliance on graphs and other diagrammatic material, makes mathematical concepts hard to convey to blind students. This is true not just in the developing world but also globally. Blind students are typically unable to pursue interests in the sciences (Ediyanto and Kawai, 2019; Maguvhe, 2015), partly due to the need for a firm grounding in mathematics. Nemeth Code, abacus, and Braille writers have been used traditionally for blind children (Brawand and Johnson, 2016) to help them do simple calculations.

As students are progressing from primary school to higher grades, they must rely on computation aids to handle more complex concepts related to algorithm flow, sequencing, geometric forms, graphing, and measurements. Recent years have seen encouraging innovations, although they are yet to make their way to widespread deployment. For instance, process-driven mathematics, based on audio methods of instruction, has been developed for students who are no longer able to use traditional low-vision tools like Braille and Nemeth Code (Gulley et al., 2017). Maćkowski et al. (2018) have developed a multimedia-enabled platform for interactive learning of mathematics by the blind. (Beal and Rosenblum, 2018) found that applications on tablet computers are motivating ways to teach mathematics concepts, eliciting greater engagement from students relative to traditional literacy methods. Despite these positive recent developments, it is fair to say that the current state of math education for the blind leaves much to be desired, as is evident from the conspicuous absence of blind students in math and science programs in schools and colleges. It is not clear precisely when in the schooling trajectory this push away from mathematics commences. Do blind children perform on par with their sighted counterparts until the middle grades, before specialized symbols and equations begin to be used heavily? Answering this question requires appropriate diagnostic tests that objectively measure students' math preparedness at different grade levels. However, designing such diagnostic tests has been a challenge as many concepts require visualization.

With this background, the two specific aims of the work we report here were:

- The design of diagnostic tests for a range of grades that could be administered to Prakash children over the telephone without the need for graphical communication
- The collection and analysis of data on diagnostic tests

7.4 METHODS

7.4.1 *Participants*

The children who participated in this study were all identified and treated as part of Project Prakash during the past 12 years. Initial identification took place at outreach eye screening camps in various states of north India. All children were diagnosed as having congenital bilateral cataracts. We contacted 75 Prakash patients (47 males, and 28 females). Of these, 17 were excluded since they were older than 20 years, and 4 had to opt out due to other obligations. Thus, our final participant pool comprised 54 children (32 males, and 22 females) ranging in age from 5 to 19 years (Figure 7.2a). There was no statistically significant difference in the ages of the male group relative to the female group (Male group: age range = $[5, 19]$, mean $= 14.5$; Female group: age range = $[8, 19]$, mean $= 13$; $t(52) = -1.6$, $p = 0.12$ in two-sample t-test). 48 children were provided free surgical treatment at Project Prakash's medical partner institution, Dr. Shroff's Charity Eye Hospital, New Delhi. Surgeries for the remaining six were delayed due to the pandemic but are expected to be conducted in the spring of 2022. The Prakash team stayed in touch with the children to monitor their visual status as well as their progress on social and educational dimensions. For the present study, the Prakash team commenced re-contacting these children in August 2021. We collected information about their general health and present educational status, i.e., whether they were enrolled in a school and, if so, in what grade.

7.4.1.1 *Current enrollment status and grade level*

Out of the 54 participants, 50 were enrolled in a school. 4 children were not enrolled in a school at the time of the study but had passed different grades (one had passed 12th grade, two had passed 10th grade, and one had passed 8th grade). Figure 7.2b shows the distribution of children across grades.

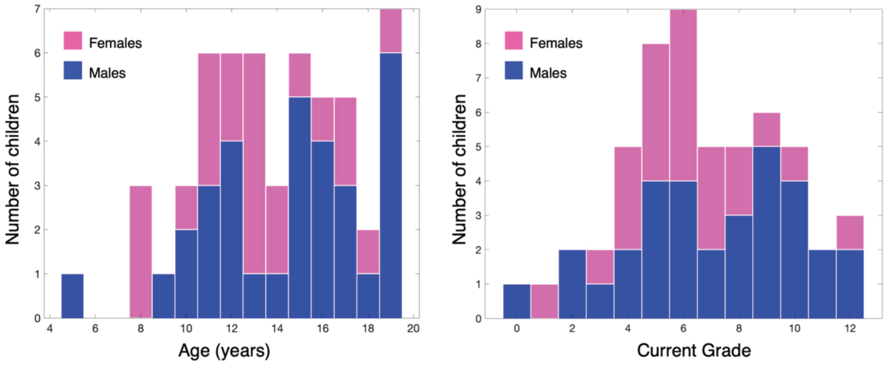


Figure 7.2: (a) Distribution of ages across our participant group, (b) The grades that students were enrolled in at the time of the study.

7.4.2 *Diagnostic assessments*

Based on the guidelines specified by NCERT and the Central Board for Secondary Education (CBSE), an Indian governmental body tasked with defining school curricula for the entire country, we formulated lesson plans and questions appropriate for grades one through five. For each grade, we identified the key topics the students are expected to understand (for example, first-grade topics included basic arithmetic operations, word problems, and applications such as time durations and carrying out a change with coin denominations). For each topic, we generated a series of questions spanning a range of difficulty levels (governed in part by the magnitude of the numbers involved in a problem; 3 + 4 being more accessible than 23 + 61) and ensured that they did not require any pictorial components like graphs or diagrams. From this pool of questions, we created two versions of each assessment, a short test for the initial assessment and a more extended test if gathering more information was deemed necessary by the proctor (as described in the next section). In developing these tests, we solicited input from two external elementary school educators at Florida Ruffin Ridley School (Brookline, MA). In the final step, a speaker fluent in English and Hindi translated the assessment material to Hindi allowing for it to be administered to children in north India. These steps were adopted for generating assessments for grades one through five.

7.4.3 *Administration of the diagnostic tests*

Tests were administered telephonically by six Prakash team members fluent in Hindi. Children were intimated about the test a day before to ensure they were available and prepared. Most questions in the assessments were designed to require only mental computations. Still, the children were free to use any writing or Braille material they wished if doing so facilitated the calculations. Assessment of each

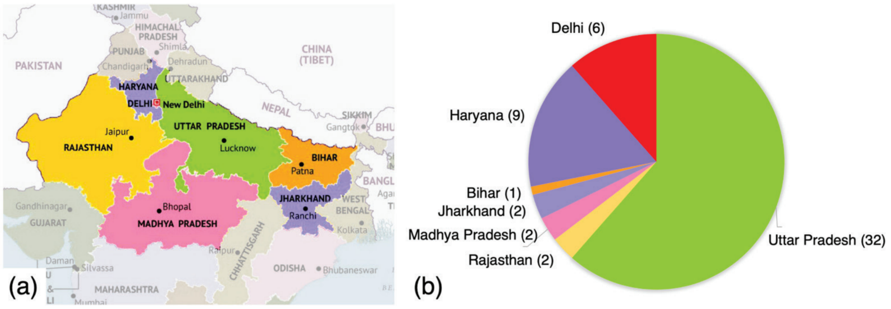


Figure 7.3: Geographical distribution of the children participating in the study. (a) Our participants came from seven north/central Indian states. (b) Number of participants from each state.

child started at grade level 1. The shorter test was administered before proceeding to the longer one. The more extended test was rendered superfluous and was not administered if the participant could not answer more than half of the questions on the short assessment. If the child scored more than 80% on the long assessment of a given level, testing moved on to the next level. The last assessment level for which the child did not reach the passing score was taken to be his/her current proficiency. Each level's assessment took about 20-30 minutes to administer over the phone.

7.5 RESULTS

7.5.1 *Demographics*

The children belonged to 7 states in north/central India. The distribution is shown in figure 7.3.

7.5.1.1 *Socio-economic status*

All children in the participant pool came from families with household incomes lower than $150 per month. Since each household comprised three or more members, their income places them below the International Poverty Line (indicating 'absolute' or 'extreme' poverty), as defined by the World Bank (2016). None of the parents had had schooling beyond fifth grade, and most were illiterate. They worked as laborers on construction sites or farms. None of the children had access to personal computers or broadband internet.

7.5.2 *Results of assessments*

7.5.2.1 *Mode of written communication*

Before their treatment, all children were enrolled in schools for the blind, where the mode of instruction was exclusively Braille. Following

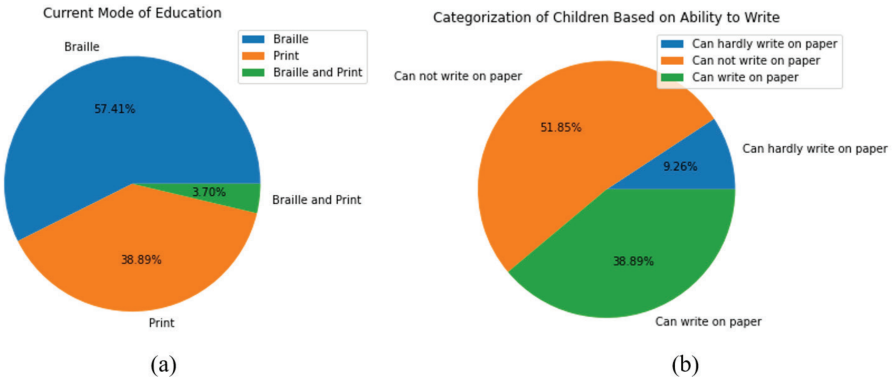


Figure 7.4: (a) Distribution of children based on their current mode of education. (b) Categorization of children based on their ability to write on paper.

surgery, despite gaining adequate vision to work with enlarged text, only a minor of the children have shifted from Braille to print (figure 7.4a). This is because many children must continue in blind schools due to the reticence of regular schools to offer them admission.

Results on the non-Braille reading/writing ability of these children follow the general pattern depicted in figure 7.4a. As shown in figure 7.4b, 51.85 % of the children are unable to write on paper, 9.26 % have a very rudimentary ability to write, and 38.89 % can comfortably write on paper (this is the group that had transitioned to print as the mode of education).

7.5.2.2 *Comparison between current and assessed grade*

Figure 7.5a compares the recommended and current education levels for all participants. Figure 7.5b plots grade levels (in which the child is enrolled, as well as what the assessments reveal) across ages and separated by gender. Strikingly, most students performed well below their putative grade levels. This was especially notable in the case of children studying in high school, who were found to have a level of mathematical readiness below grade 4.

7.6 DISCUSSION

To summarize, we contacted 54 Prakash children and administered diagnostic tests that we designed for grades 1 through 5. The tests were presented in the children's vernacular to ensure that language was not a limiting factor in their performance. The data reveal that most children have math skills at or below grade level 3. This is noteworthy especially given that several of the children are in their mid or late teens and are nominally enrolled in advanced grades in their schools. Figure 7.6 shows the distribution of recommended grade levels for the participating children. This distribution is in marked contrast to

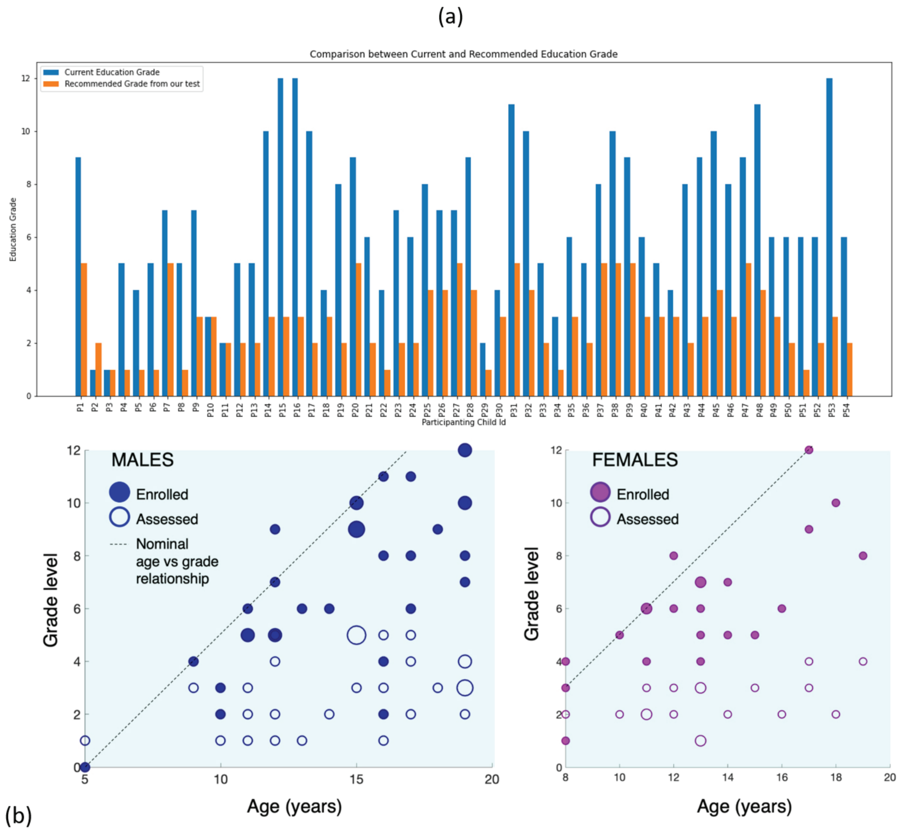


Figure 7.5: (a) Comparison between the current and recommended grade. (b) Scatterplots of enrolled and assessed math proficiency across ages for male and female Prakash children.

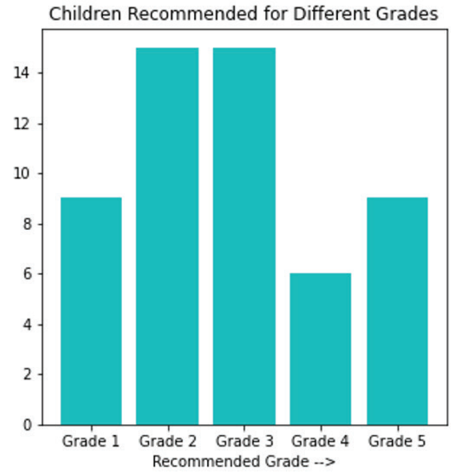


Figure 7.6: Children recommended for different grades.

the one shown in figure 7.2b. It is clear from this discrepancy that the children's actual level of proficiency lags well behind what their age or grade would suggest. This alarming situation reveals the deficiencies of the current education programs the children are enrolled in and makes a compelling case for more effective educational interventions.

Further studies are needed to identify the factors underlying the problems associated with the current schooling systems for the blind. However, some likely candidates include the poor training of special education teachers and the lack of oversight of schools for the blind. The policy of automatic promotion of children from one grade to the next, even if they do not meet scholastic readiness standards for the higher grade (Right to Education Act, 2018, Government of India), likely contributes to the kind of situation we have witnessed in our results, with putative 12th graders performing below the level of a 4th grader. This policy needs to be rethought.

A caveat that needs to be kept in mind while interpreting these findings is that the diagnostic assessments only addressed math proficiency, which even though important, is only one element of the whole curriculum. It is quite possible that the results would be different depending on the subject assessed. Further studies are needed to comprehensively determine the status of any discrepancies between expected and actual levels of scholastic preparation across other subjects drawn from languages, humanities, and sciences.

7.6.1 *Considerations for designing a new education program*

The exercise of administering diagnostic tests to the Prakash children and analyzing the data helps guide our thinking about designing a new education program that can address some of the problems with the current system. The content and organization of this new program must be considered carefully since it has to satisfy multiple constraints, including conformity with state curricula, compatibility with the children's lingering visual deficits, and the necessity to allow for dynamic, self-paced progression through the program. Additionally, the parents would need to be convinced that the educational intervention is in the child's best interests to improve his/her prospects of eventually finding a job that can provide them financial independence.

It appears likely that the education program would need to be digitally administered, given the inadequate physical schooling infrastructure. Such a program would have to contend with technological challenges. Although basic phone connectivity has reached most remote places in India, there are still severe bandwidth and stability issues. To be able to provide education remotely, we, therefore, cannot rely exclusively on synchronous online instruction but need to develop more appropriate solutions. One option would be to preload personal digital tablets with specifically designed scholastic material. Several studies, such as Dorouka et al. (2020), Hubber et al. (2016), and Mulet et al. (2019) have assessed the utility of tablet-based interventions in primary education, generally suggesting that tablets are helpful as they are easy to use, enhance interest in the task, and allow students to be independent, thereby encouraging self-education and self-confidence. We believe that combining the tablet-based delivery of scholastic material with regular telephonic check-ins could greatly aid in bridging the educational gap these children face and help bring them to desired academic proficiency levels. An important point to consider in the implementation of such a tablet-based educational program, however, is the sub-par visual status of the students (primarily due to lower than 20/20 acuity (Ganesh et al., 2014), and reduced contrast sensitivity (Kalia et al., 2017). While tablets are a visual aid, we need to assess how they can be best utilized for low-vision patients and understand what educational apps and content would most benefit the children.

Another aspect worth considering in the design of our education program is that many concepts in mathematics require exposure to various shapes. Visually impaired children are often deficient in these experiences. To remedy this in part, physical aids could synergistically enhance the process of teaching to the visually impaired. Manipulable objects in many forms may assist in learning mathematical concepts and increase comprehension accuracy in visually impaired students (Brawand and Johnson, 2016). Providing physical aids would significantly facilitate the teaching of concepts in geometry.

Finally, the question arises whether, considering the lower-than-normal visual status of many of the Prakash children, including tests of visual function on the digital tablets would be helpful to complement the educational content provided. We believe that the answer is in the affirmative. Assessing Prakash children's visual status and development over time would allow for a more continuous evaluation of their visual health. Furthermore, it could inform the visual parameters (such as the size, contrast, and color of fonts or other items shown on the tablets) with which the educational contents are being presented. Including such visual assessments may not only serve an essential educational and clinical purpose but could also provide an unprecedented window into the study of visual development by enabling a longitudinally dense tracking of several aspects of visual development, such as visual acuity or color sensitivity profiles.

To conclude, the data we have gathered reveal marked discrepancies between the actual and expected levels of scholastic math preparation in children who have transitioned from blindness to sight. This points to the urgent need to introduce more effective educational approaches for children with visual impairments. We have outlined some considerations relevant for designing a new, potentially digital, education program. Success in thisundertaking will be impactful not only for improving educational prospects for children with visual impairments but also for the immensely large population of 'under-schooled' children, whose educational preparation is much below age-appropriate levels.

ACKNOWLEDGEMENTS

This study was supported by a grant from MIT J-WEL. Project Prakash is supported by the National Eye Institute (NIH) via grant R01EY020517, the Sikand Foundation, the Nick Simons Foundation, and the Halis Foundation. The authors wish to thank all the Prakash children and their families who participated in this study.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

Beal, Carole R. and L. Penny Rosenblum (2018). "Evaluation of the Effectiveness of a Tablet Computer Application (App) in Helping Students with Visual Impairments Solve Mathematics Problems." In: *Journal of Visual Impairment & Blindness* 112.1, pp. 5–19.

Bell, Edward C. and Arielle Silverman (2019). "Access to Math and Science Content for Youth Who Are Blind or Visually Impaired." In: *Journal of Blindness Innovation and Research*.

Brawand, Anne C and Nicole M Johnson (2016). "Effective Methods for Delivering Mathematics Instruction to Students with Visual Impairments." In: *Journal of Blindness Innovation and Research* 6.1.

Dorouka, Pandora, Stamatis Papadakis, and Michail Kalogiannakis (2020). "Tablets and apps for promoting robotics, mathematics, STEM education and literacy in early childhood education." In: *Int. J. Mob. Learn. Organisation* 14, pp. 255–274.

Ediyanto and N Kawai (2019). "Science Learning for Students with Visually Impaired: A Literature Review." In: *Journal of Physics: Conference Series* 1227.1, p. 012035.

Gandhi, Tapan K, Amy Kalia Singh, Piyush Swami, Suma Ganesh, and Pawan Sinha (2017). "Emergence of categorical face perception after extended early-onset blindness." In: *Proceedings of the National Academy of Sciences* 114.23, pp. 6139–6143.

Ganesh, Suma, Priyanka Arora, Sumita Sethi, Tapan K Gandhi, Amy Kalia, Garga Chatterjee, and Pawan Sinha (2014). "Results of late surgical intervention in children with early-onset bilateral cataracts." In: *British Journal of Ophthalmology* 98.10, pp. 1424–1428.

Gulley, Ann P, Luke A Smith, Jordan A Price, Logan C Prickett, and Matthew F Ragland (2017). "Process-driven math: An auditory method of mathematics instruction and assessment for students who are blind or have low vision." In: *Journal of visual impairment & blindness* 111.5, pp. 465–471.

Gupta, Priti, Pragya Shah, Sharon Gilad Gutnick, Marin Vogelsang, Lukas Vogelsang, Kashish Tiwari, Tapan Gandhi, Suma Ganesh, and Pawan Sinha (2022). "Development of visual memory capacity following early-onset and extended blindness." In: *Psychological Science* 33.6, pp. 847–858.

Hubber, Paula J, Laura A Outhwaite, Antonie Chigeda, Simon McGrath, Jeremy Hodgen, and Nicola J Pitchford (2016). "Should touch screen tablets be used to improve educational outcomes in primary school children in developing countries?" In: *Frontiers in Psychology* 7, p. 839.

Kalia, Amy, Tapan Gandhi, Garga Chatterjee, Piyush Swami, Harvendra Dhillon, Shakeela Bi, Naval Chauhan, Shantanu Das Gupta, Preeti Sharma, Saahil Sood, et al. (2017). "Assessing the impact of a program for late surgical intervention in early-blind children." In: *Public Health* 146, pp. 15–23.

Maćkowski, Michał, Piotr Brzoza, Marek Żabka, and Dominik Spinczyk (2018). "Multimedia platform for mathematics' interactive learning accessible to blind people." In: *Multimedia Tools and Applications* 77, pp. 6191–6208.

Maguvhe, Mbulaheni (2015). "Teaching science and mathematics to students with visual impairments: Reflections of a visually impaired technician." In: *African journal of disability* 4.1, pp. 1–6.

Mulet, Julie, Cécile Van De Leemput, and Franck Amadieu (2019). "A critical literature review of perceptions of tablets for learning in primary and secondary schools." In: *Educational Psychology Review* 31, pp. 631–662.

Ostrovsky, Yuri, Ethan Meyers, Suma Ganesh, Umang Mathur, and Pawan Sinha (2009). "Visual parsing after recovery from blindness." In: *Psychological Science* 20.12, pp. 1484–1491.

Rahi, JS, S Sripathi, CE Gilbert, and Allen Foster (1995). "Childhood blindness in India: causes in 1318 blind school students in nine states." In: *Eye* 9.5, pp. 545–550.

Rohwerder, Brigitte (2018). *Disability stigma in developing countries*.

Sinha, Pawan (2013). "Once blind and now they see." In: *Scientific American* 309.1, pp. 48–55.

Sreekanth, Y (2016). *Eighth All India School Education Survey - A concise report*. Tech. rep. New Delhi: National Council of Educational Research and Training.

World Bank (2016). *Vision impairment and blindness, Fact Sheet N^{o}282*. https://www.worldbank.org/en/news/feature/2016/01/13/principles-and-practice-in-measuring-global-poverty, (access:2022/12/21).

Part III

DISCUSSION

8 DISCUSSION

8.1 SUMMARY OF FINDINGS

In Chapters 2-4, I have, across a few stimulus dimensions, presented computational and, in part, experimental examinations of the consequences of initially degraded sensory experience for the developing perceptual system. These consequences have been further described and discussed in Chapter 5. In Chapters 6&7, I have reported additional contributions to tests of the visual memory capacity and scholastic performance of Prakash individuals. Finally, in Appendix A&B, I have reported additional computational contributions to normal development and normal perception in adults. I will summarize the main findings of these studies below, before discussing the key insights that emerge from them in greater detail.

In Chapter 2, I have reported computational tests of the AID hypothesis in the domain of prenatal hearing, in an attempt to examine the consequences of commencing auditory development with strongly low-pass-filtered sounds in the intrauterine environment. These tests have revealed that deep networks trained with a developmentally-inspired temporal progression of inputs (transitioning from initially low-pass-filtered inputs to full-frequency ones later on) differed from networks trained on exclusively full-frequency inputs in two crucial regards. First, biomimetic training resulted in the formation of temporally extended, stable receptive field structures. Second, such training led to markedly more generalized performance on emotion recognition – a task relying on the analysis of information across extended time intervals. Interestingly, these computational results not only attest to the functional significance of typical developmental trajectories but also help account for specific deficits in the analysis of low-frequency sound structure that have previously been reported in prematurely born babies. The latter are particularly relevant here as their periods of exposure to initially low-pass-filtered inputs were markedly shortened, thereby superficially resembling the non-biomimetic training regimen comprised of full-frequency inputs right from the start. Together, these observations provide support for the AID proposal in the domain of prenatal hearing.

In Chapter 3, utilizing a similar computational approach, I have extended the examination of the role of initial degradations to the domain of color vision, inspired by the presence of poor color sensitivity immediately after birth. The computational results compellingly demonstrated that training with a developmentally inspired progression of color-degraded to color-rich inputs critically prevented the emergence of processing mechanisms exhibiting an overly strong focus on chromatic details. This has been achieved by instantiating representations that are rather tuned to luminance and shape information, resulting in stable behavioral classification, even upon color removal or color shifts. Such mechanisms, in turn, provide a candidate account for how humans come to acquire their remarkable ability to recognize objects under changing color conditions. The computational results presented were further complemented by experimental data on Prakash individuals following their sight-restoring surgeries. These experiments revealed that while generally proficient in object recognition, Prakash individuals, as opposed to normally-sighted controls, perform very poorly when asked to name objects that are presented in grayscale. In agreement with the AID hypothesis, this could, as a second experiment confirmed, potentially be accounted for by their color vision system being fully matured already days after the surgeries. As such, the empirical and computational results together point to the potential significance of commencing vision with initially color-degraded inputs, and the detrimental consequences of missing out on this period marked by initial degradations.

In Chapter 4, I conducted computational explorations of the consequences of joint progressions in sensory development. The motivation for these investigations was that, in the experience of a newborn, the development of several sensory dimensions (such as visual acuity and color sensitivity) is not independent but, instead, occurs together. I have computationally tested whether this joint progression may help account for an important organizational principle of the visual system: the division into the magnocellular and the parvocellular pathway, which is characterized by differences in spatial frequency sensitivity and color sensitivity. This investigation was particularly driven by the consideration that visual experience at birth is marked by low acuity and color vision and that receptive fields established early in development could consequently jointly come to encode such magnocellular-like attributes. Later, when higher acuity and color information are available, parvocellular-like receptive fields could be established. My computational simulations support this hypothesis, both in terms of the resulting receptive field structures and their behavioral correlates. Further, training with the joint biomimetic training regimen has led to the emergence of a more human-like global shape bias in classification, thereby also providing inspiration for the design of advanced training procedures for deep neural networks.

Together, the studies reported in Chapters 2-4 attest to the potential adaptive significance of the development of perceptual proficiencies from poor to proficient. This significance has been further elaborated on in Chapter 5.

In Chapter 6, I have included additional computational contributions to empirical work that aimed to assess the visual memory capacity of Prakash children. The experimental results revealed that the visual memory capacity of the Prakash group is poor immediately post-surgically but improves gradually over the months to follow, almost reaching the levels of normally-sighted controls one year after the surgeries. The computational work I contributed to this work furthermore motivates that the longitudinal improvements in memory performance may not necessarily be memory-specific but could potentially also be attributed to the progressive representational elaboration of the perceptual system, critically driven by experience. The latter would have the consequence of rendering different images in a set of images more discriminable and thereby facilitate their memorization.

In Chapter 7, I included non-computational contributions to highly collaborative work on examining the educational profile of Prakash children. In brief, the data highlighted the marked deficits of Prakash individuals in terms of their scholastic performance and overall educational outlook. These results render the mission of improving the educational opportunities for the late-sighted, as well as many other groups of children, an especially crucial and urgent one.

In Appendix A, I have included additional computational contributions to an empirical assessment of the normal development of visual memory capacity in children and adults. The experimental results revealed performance differences between the two populations, indicating that certain aspects of visual memory development last until late childhood. In addition, the data revealed higher memory performance for 'meaningful', natural images, as opposed to images of abstract art. The computational work I contributed to this project focused on examining whether the performance differences between the two sets of image classes may be explained by experience with real-world natural imagery. While this is a possibility, the computational results have primarily highlighted the role of experience and have also illustrated important differences between the human and artificial visual system.

Finally, in Appendix B, I have included additional computational contributions to face recognition at a distance. The experimental component of the study revealed that the recognition of faces at a distance differs markedly from the recognition of nearby faces, with the former explicitly relying on the relationships between internal facial features and external head contours. Computationally, I contributed to this study by examining whether processing mechanisms relying on such relationships could be learned through natural experience with faces. These computational results, while providing some support for this proposal, also reveal crucial differences between the processing of the human visual system on the one hand and deep networks on the other.

Across the above studies, several observations emerge. I will discuss these in greater detail below, particularly concerning the AID hypothesis and the post-surgical status of late-sighted individuals.

8.2 ON THE ROLE OF INITIAL SENSORY DEGRADATIONS AND EARLY PERCEPTUAL EXPERIENCE

The results presented in Chapters 2 and 3, as well as past results on visual acuity reviewed in Chapter 5, are remarkably consistent across computational and experimental investigations and across the different perceptual dimensions tested. Together, these studies provide support for the AID hypothesis, positing that initial limitations as part of developmental progressions may be of assistance, rather than represent hurdles, for the acquisition of perceptual proficiencies.

Considering the consistent methodologies that were used for the work presented in Chapters 2 and 3, as well as for the past work on visual acuity reviewed in Chapter 5, fairly detailed and direct comparisons can be made across studies. An entirely consistent finding is that training with biomimetic regimens leads to markedly better generalization performance than does training on full-quality inputs throughout the entire training process. Some across-study differences can be observed, however, in the other two regimens: training exclusively on degraded inputs and training using an inverse-biomimetic regimen. While, in the studies on visual acuity and prenatal hearing, training exclusively on degraded inputs does not induce broad generalization, it does lead to fairly good generalization performance in the study on color sensitivity. While the definite origin of these differences is yet to be determined, one possible account could be based on the observation that only very specific aspects of object recognition are likely to directly benefit from the use of chromatic cues (such as the fine-grained discrimination between certain food classes). Thus, the differential importance of color on the one hand and high spatial/temporal frequencies on the other could potentially account for the observed differences in generalization following training on exclusively degraded inputs.

A second observation that emerges from a detailed cross-study comparison is that, while the inverse-biomimetic regimen led to the worst generalization performance in the studies of visual acuity and color sensitivity, it was the second-best in the study on prenatal hearing. As the former two are both visual and the latter is auditory, differences between the network architectures or the training data may account for these differences. Thus, follow-up investigations with more diverse network architecture, training settings, and datasets could help reveal the origin of this difference.

A few general caveats should be kept in mind when interpreting tests of the AID hypothesis. First, as already pointed out in Chapter

5, the observed benefits resulting from initially degraded perceptual experience are not definite proof that initial degradations are 'intended' to yield such benefits. In other words, we cannot exclude the possibility that the results could be epiphenomenal nevertheless. More practically, it is also important to remember that the computational results have been generated with deep neural networks, which, as described in Section 1.4, are not perfect models of the human perceptual system. Further, there are many different architectures, training sets, and network parameters across which computational results are yet to be generated (see Section 8.4.).

While the above caveats are important to keep in mind, they should not distract from the consistency of the support that the studies described in this thesis have presented for the AID hypothesis. In addition, as detailed in Chapter 4, the joint developmental progressions of visual acuity and color sensitivity have provided a candidate account for the origin of the division into parvo- and magnocellular pathways. Thus, both the specific approach of testing the AID hypothesis and the more general approach of incorporating aspects of normal development into computational modeling emerge as powerful research avenues.

8.3 THE STATUS OF LATE-SIGHTED INDIVIDUALS FOLLOWING SIGHT-RESTORING SURGERIES

The studies presented in Chapters 3, 6, and 7 have meaningfully added to our understanding of the post-surgical development of late-sighted individuals.

The study on visual memory reported in Chapter 6 has revealed that, while not fully reaching the visual memory capacity of normally-sighted controls, one year following the sight-restoring surgeries, the Prakash group is remarkably close to the controls. This observation allows us to add visual memory capacity to the set of skills that are fairly resilient to extended, early-onset visual deprivation. This, in turn, has important implications for our conceptualization of plasticity late in life, which speaks against the notion of strict critical periods of development. In addition, the computational component of this work added plausibility to the idea that the longitudinal improvements in visual memory capacity following the surgeries could potentially be driven by increases in representational elaboration rather than memory capacity per se. This provides an interesting starting point for future investigations.

The study on the usage of color cues, as described in Chapter 3, has added to our current understanding of recovery from congenital blindness in two ways. First, the observation that (almost) immediately following the surgeries, Prakash children had color sensitivities that were comparable to controls, indicates complete resilience of this perceptual proficiency to visual deprivation. However, the study also revealed that Prakash individuals had difficulty recognizing objects that were presented in grayscale, adding to the complex landscape of recovery from visual deprivation. As described earlier, the two results taken together have provided further empirical support for the AID account.

In light of this observed deficit, it is worth highlighting an interesting difference between the conceptualization of critical periods on the one hand and the AID account on the other. Specifically, the reduced robustness of object recognition to chromatic variations described in Chapter 3, as well as the observed deficits in the holistic processing of faces (as described in the acuity study reviewed in Chapter 5) could, in principle, be accounted for by a critical-period account that is specific to certain aspects of vision. Such an account would critically rely on the assumption that in the absence of appropriate visual inputs early in life, certain perceptual skills can no longer be acquired due to too low cortical plasticity. This account differs from the AID account on a theoretical level. Specifically, the latter posits that it is not the absence of inputs early in life but the specific quality of inputs following delayed sight onset that is responsible for the subsequently observed perceptual deficits. As an important consequence, the AID account would predict that changes to the stimulus quality immediately post-surgically could potentially enable the late-sighted to acquire such proficiencies. To the contrary, this hope would be absent in an account based on critical periods. The more “optimistic” view resulting from the AID proposal could, in principle, be tested, as is further detailed in Section 8.4.

Finally, it is crucial to note that while some aspects of visual proficiency do not reach the level of normally-sighted individuals, the greatest deficits are observed outside the domain of vision. Specifically, as described in Chapter 7, the educational status of Prakash children, even if enrolled in schools, falls tragically short of what it could be, and highlights an urgent humanitarian need outside of the perceptual domain.

8.4 FUTURE DIRECTIONS

There are several ways in which the work presented in this thesis could be further expanded in the future. I highlight below a few particularly noteworthy and promising research avenues, split into computational and experimental investigations.

8.4.1 *Computational modeling*

First, the generalizability of the findings that have resulted from computational tests of the AID hypothesis could be examined further.

For instance, future computational work could draw on additional network architectures, such as those that also include recurrent connections (e.g., Kubilius et al., 2019). In addition, the results could be compared systematically across datasets of different object classes, such as faces vs. non-face objects. Also, more ecologically-motivated datasets could be utilized for training, such as the Ecoset database (Mehrer et al., 2021), or even datasets that were created to specifically mirror the kinds of stimuli to which newborns are typically exposed.

Further, additional analyses and comparisons could be carried out to help evaluate whether the training with developmentally-inspired regimens could result in better models of the human perceptual system or more robust machine learning models for the field of engineering. To evaluate the former, systematic examinations of the representational similarity between biomimetically-trained networks and neural as well as behavioral recordings could prove to be especially impactful. To evaluate the latter, it would be important to compare the robustness of biomimetically trained models with models that were not trained developmentally but rely on heavy data augmentation.

In addition to testing the generality and the potential implications of the results shown for the previously examined perceptual dimensions, future tests of the AID hypothesis could also be extended to additional aspects of development. One such aspect was already hinted at in Chapter 5: the temporal evolution of the use of linguistic labels, as motivated by the observation that parents are more likely to describe objects to their young children using basic, rather than subordinate, labels. In addition to further perceptual dimensions, future examinations could also study the role of changes in the neural architecture that occur throughout the developmental timeline – for instance, those related to the development of feed-forward vs. feedback connections (e.g., Berezovskii et al., 2011).

Finally, a particularly interesting addition to the computational simulations reported would be the incorporation of the time dimension. Practically, this could be achieved by working, for instance, with 3D convolutional neural networks that are processing videos instead of static images. Such extension could help move the investigation of adaptive perceptual development from the purely spatial to the spatiotemporal domain. The incorporation of the temporal dimension could thereby prove especially crucial for the work reported on the parvo/magno pathways. Specifically, one could examine whether not only the spatial but also the temporal characteristics of the parvo/magno distinction could be accounted for by developmental experience.

8.4.2 *Work with late-sighted individuals*

Similar to the importance of incorporating the time dimension into deep neural networks, carrying out empirical examinations of Prakash individuals' temporal processing abilities could provide additional insight into atypical perceptual development. As past examinations of the resilience of visual proficiencies to deprivation have primarily focused on spatial aspects of vision, it would be important to complement these assessments with temporal proficiencies. This could, among others, include probing the ability of the late-sighted to perceive simultaneous vs. non-simultaneous events and to detect temporal correlations in sensory streams.

In addition, to further examine the potential consequences of late-sighted individuals commencing vision with abnormally high initial visual acuity and color sensitivity, further perceptual tests could be carried out. For instance, in the specific context of the AID hypothesis, one could examine the decision biases of the late-sighted in classifying images on texture vs. shape information, or assess the performance on additional tasks that rely on extended spatial integration, such as contour integration.

In addition to these indirect types of empirical validation, the most definite evidence could, at least theoretically, be obtained in the context of rehabilitative interventions. For instance, this could include the design of specifically controlled auditory sound environments for neonatal ICUs (to provide auditory signals more similar to the low-pass filtering in the intrauterine environment) or the design of glasses that artificially limit high spatial frequencies and color information immediately following the sight-restoring surgeries of Prakash children. I am not able to assess how realistic such interventions are to be implemented in the future or how successful they would be. However, it is essential to highlight that, if successful, they would provide the most direct form of empirical validation for the AID hypothesis and would result in direct clinical benefits for the late-sighted.

Finally, continuing on the applied front, the work presented in Chapter 7 has highlighted a remarkable humanitarian need for improving the educational perspectives of late-children children (in addition to other populations of undereducated children). In the future, the design of a 'bridge course' could potentially allow such individuals to catch up on several grades in a relatively short amount of time, to be able to be reintegrated into the school system. If successful, even if only partially, this effort would represent another, broader, 'Prakash opportunity' that would even exceed the domain of perception.

8.5 CONCLUSION

To conclude, in this thesis, I have presented computational tests of the AID hypothesis, positing that experience with initially degraded inputs may help, instead of hinder, the acquisition of perceptual proficiencies. Together, these investigations have provided converging evidence in favor of the hypothesis. This has important implications for several scientific domains. For the field of developmental science, this work can contribute to our understanding of the potential functional significance of normal developmental trajectories. For the field of artificial intelligence, biomimetic regimens could inspire the design of more robust training procedures for deep neural networks. More generally, the approach chosen is illustrative of how human studies can lead to advances in computational modeling. Finally, from the clinical perspective, the AID proposal has the potential to help guide prognoses following sight-restoring surgeries and motivate the design of rehabilitative interventions in the future. Some of the work presented in this thesis has also featured empirical studies of late-sighted individuals. These studies, in addition to providing additional support for the AID account, have also advanced our understanding of the resilience of certain visual skills to early-onset, extended visual deprivation. Together, these studies have motivated several follow-up investigations, both on the computational and the experimental front, and also highlighted the humanitarian need for improving the educational perspectives of the late-sighted and other underprivileged children.

REFERENCES

Berezovskii, Vladimir K, Jonathan J Nassi, and Richard T Born (2011). "Segregation of feedforward and feedback projections in mouse visual cortex." In: *Journal of Comparative Neurology* 519.18, pp. 3672–3683.

Kubilius, Jonas, Martin Schrimpf, Kohitij Kar, Rishi Rajalingham, Ha Hong, Najib Majaj, Elias Issa, Pouya Bashivan, Jonathan Prescott-Roy, Kailyn Schmidt, et al. (2019). "Brain-like object recognition with high-performing shallow recurrent ANNs." In: *Advances in neural information processing systems* 32.

Mehrer, Johannes, Courtney J Spoerer, Emer C Jones, Nikolaus Kriegeskorte, and Tim C Kietzmann (2021). "An ecologically motivated image dataset for deep learning yields better models of human vision." In: *Proceedings of the National Academy of Sciences* 118.8, e2011417118.

Part IV

APPENDIX

A THE INFLUENCE OF SEMANTICS ON LONG-TERM VISUAL MEMORY CAPACITY IN CHILDREN AND ADULTS

Content from

Gupta, P., Shah, P., Gilad-Gutnick, S., **Vogelsang, M.**, Vogelsang, L., & Sinha, P. (Submitted). "The influence of semantics on visual memory capacity in children and adults". **Under revision** at British Journal of Developmental Psychology (initial submission: December 2022)

A.1 ABSTRACT

Human visual memory capacity has a rapid developmental progression. Here we examine whether image semantics modulate this progression. We assessed the performance of children (6-14 years) and young adults (19-36 years) on a visual memory task using real-world or meaningful and abstract image sets, which were matched in low-level image attributes. For real images, we find comparable performance across the two age groups, consistent with previously reported results. However, for abstract images, we find a clear age-related difference indicating greater reliance of children's memory processes on semantics, indicating that strategies for encoding abstract patterns keep improving even into late childhood. We complemented these studies with computational experiments designed to examine the role of increasing experience with real-world images on real and abstract image encoding, to examine whether the observed age-related differences as well as the general privilege of real over abstract images can emerge directly through experience with meaningful images. Our results provide support for this possibility, and set the stage for a finer-grained investigation of the timeline along which children's memory capacity for abstract images reaches adult levels.

A.2 KEYWORDS

Visual memory capacity, picture memory, image semantics, real images, abstract images

A.3 INTRODUCTION

Imagine a morning spent visiting an art museum. You and your eight-year-old niece stroll through halls with paintings from realists such as Winslow Homer, Gustave Courbet, and Andrew Wyeth. Other exhibits show abstract paintings from artists such as Wassily Kandinsky, Piet Mondrian, and Willem de Kooning. You linger at each painting for a few seconds. At home, later in the day, you browse through a large compendium of art, which contains pictures of numerous paintings, including some that you saw in the art museum. Will your ability to tell whether a picture depicts a painting you saw earlier differ depending on whether it is a realistic painting or an abstract one? And, will a painting's meaningfulness have different impact on your and your niece's ability to remember it?

This thought experiment relates to two broad scientific questions. The first concerns how meaning influences long-term memory, which has been an active and fruitful avenue of research and provides the backdrop for the second, and main, question we are considering in this paper: do semantic cues differ in their significance for children on the one hand and adults on the other?

How meaning influences memory has been examined in domains such as word recall (Bower et al., 1969; Calkins, 1898; Collins and Quillian, 1969; Dooling and Lachman, 1971; Ernest and Paivio, 1971; Paivio, 1971; Sasson and Fraisse, 1972; Tulving et al., 1972), where it was found that meaningfulness appears to facilitate memory (Pezdek, 1977; Slamecka, 1985). This issue has also been studied in the visual domain for both short-term and long-term memory (Asp et al., 2021; Bellhouse-King and Standing, 2007; Brady and Störmer, 2022; Goldstein and Chance, 1971; Konkle et al., 2010; Kouststaal et al., 2003; Madigan, 2014; Shoval et al., 2023; Standing, 1973). Several interesting results have emerged. For instance, Konkle et al. (2010) provide evidence that observers' capacity to remember visual information in long-term memory depends more on conceptual structure than on perceptual distinctiveness. They found that memory for object categories with conceptually distinctive exemplars is better and shows less interference effects as the number of exemplars increases. Shoval et al. in their recent study (Shoval et al., 2023) demonstrated that meaning not only assists visual long-term memory but is rather critical for remembering large amounts of visual information. They suggest that semantic/conceptual information acts as a 'gluing' process that is necessary for binding together independent visual features. This 'gluing' process could take place either during encoding or memory storage and could also serve as an efficient retrieval cue that facilitates accurate recognition.

Given the crucial role visual memory plays in enabling many aspects of cognitive function, it is not surprising that this resource has a rapid developmental progression. The capacity for both visual working and long-term memory increases significantly over the first year of life (Olson, 1976; Rose et al., 2001; Ross-sheehy et al., 2003; Slater, 1989) and continues to increase with age in childhood (Cowan et al., 2011; Forsberg et al., 2022; Riggs et al., 2006; Simmering, 2012; Walker et al., 1994). Children as young as four years of age spontaneously encode a high degree of visual detail in their long-term memories. They exhibit high fidelity in their visual memory capacity over a large set of items not only for basic-level categories but also for unique details and information about the position and arrangement of parts (Ferrara et al., 2017). While the topic of visual memory in infants and young children has been explored extensively, developmental literature on how semantics or meaningfulness modulates visual memory in children and how this modulation varies across different age groups is fairly sparse. Some of the studies suggest a facilitation of children's memory performance by semantics. For example, Boucher et al. (2016) reported faster and more accurate identification of concrete pictures by children as "new" or "old" compared to abstract pictures. Starr et al. (2020) also found evidence for a mnemonic benefit for familiar compared to unfamiliar objects, in children as well as adults in working memory. In another study Goujon et al. (Goujon et al., 2022) investigated, in both adults and nine-year-old children, how visual long-term memory for images is affected over time, depending on whether they were meaningful or meaningless. Participants were exposed to hundreds of meaningless and meaningful images presented once or twice for either 120 ms or 1920 ms and their memory was assessed using a recognition task either immediately after learning or after a delay of three or six weeks. They found that multiple and extended exposures, rather than meaningfulness, were crucial for retaining an image for several weeks. This pattern was observed for both adults and nine-year-old children, highlighting that although semantic information enhances the encoding and maintaining of images in long-term memory when assessed immediately, this seems not critical for long-term memory over weeks. Given the mixed nature of these findings and the scanty data on the role of semantics on long-term memory and its developmental progression, the question of how meaning modulates long-term memory as a function of age is still open.

Building on the existing literature thus far, on the role of semantics on long-term visual memory, the work outlined in this paper not only lends support to the view that image meaningfulness improves memory capacity regardless of age, but also adds to the existing knowledge in three ways. First, we have studied the developmental progression of visual long-term memory in a broader age range of children and adults. Second, we have done a systematic study of memory capacity by gradually increasing the number of items to be remembered, thereby assessing the cardinality where the capacity for meaningful and abstract images begins to differ. Third, we have complemented human data with computational work probing whether the behaviourally derived results can be explained by low-level differences in stimulus characteristics or reproduced by a computational system trained on meaningful/naturalistic imagery.

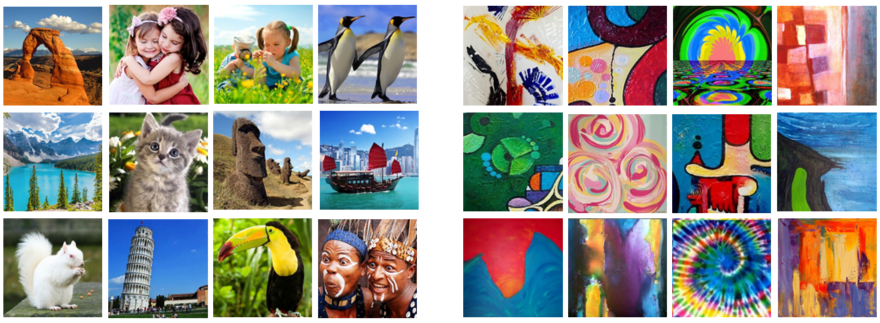


Figure A.1: Sample stimuli used in our study. The full stimulus set comprised 1200 images of real-world scenes (12 samples shown on the left) as well as 1200 images of abstract paintings (12 samples shown on the right).

A.4 HUMAN STUDIES

A.4.1 *Methods*

A.4.1.1 *Participants*

Two groups participated in this study: Group 1 comprised 20 school children (10 females; 6 to 14 years; mean: 9.6 years, SD: 2.0 years), and group 2 comprised 20 adults (10 females; 19 to 36 years; mean: 23.75 years, SD: 5.3 years). Participants were tested individually in a well-lit room on a 17-inch monitor. All participants had normal or corrected-to-normal vision and normal color vision. None had any neurological or psychiatric diagnoses or a history of visual impairment. Informed consent was taken from all participants and the study was approved by our institute's IRB.

A.4.1.2 *Stimulus set*

Stimulus set: We compiled a database of 1200 color images of real-world objects or scenes, as well as 1200 color images of abstract paintings (Figure A.1), square-cropped and scaled to the same size (256 x 256 pixels). Real images depicted a variety of natural scenes including architecture, flora, fauna, vehicles and people (Fig. A.1 right). Abstract images comprised non-representational paintings drawn from several digital art archives (Fig. A.1 left). The stimuli were presented on a 17-inch monitor using Psychtoolbox (Brainard and Vision, 1997), at a viewing distance of 40 cm. Each image subtended a visual angle of 15 degrees vertically and horizontally.

A.4.1.3 *Experimental procedure*

Our experimental design employed a multi-session 'old-new' paradigm, and each session consisted of consecutively presented encoding and

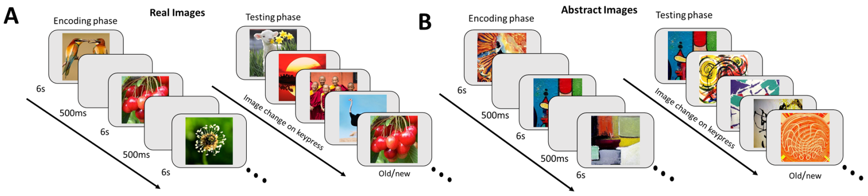


Figure A.2: Shows the schematic of the experimental methodology. The experiment was conducted in two phases: encoding and testing. During the encoding phase participants were asked to memorize multiple successively presented images, displayed for 6s each. During the testing phase, the set of previously presented training images was mixed with an equal number of novel distractors and participants were asked to verbally indicate whether a given image was previously seen 'old' or not 'new'.

test phases. During the encoding phase, participants were asked to memorize multiple successively presented images, displayed for 6s each. To sustain attention, participants were asked to rate how beautiful they found each picture to be on a scale of 1 to 5. All participants had 6s to complete the ratings. This process has been used to elicit deeper encoding which may enhance memory performance (e.g. Baddeley and Hitch, 2017; Moulin et al., 2005). For children below ten years of age, so as to not tax them too much, the rating requirement was kept optional. They could just say whether they liked the picture or not. Even though it was optional for children below ten years, all children completed the ratings for all the pictures and testing conditions, except one six-year-old child. The six-year-old child however mentioned, for all the testing conditions, if he liked or disliked the pictures. During the testing phase, the set of previously seen images was mixed with an equal number of novel distractor images. In each trial, participants were asked to verbally indicate whether a given image was previously seen or not. Each test session was conducted 2-5 minutes after memorization. To assess memory capacity, the number of images shown in the training phase was systematically increased, such that subjects were asked to remember 10, then 20, 40, and finally 80 images. Thus, each subject participated in a total of eight different experimental conditions in which they had to memorize and recognize the aforementioned four stimulus sets, for real and abstract images each. The testing order for the real and abstract conditions was randomized, in order to counter the effect of fatigue or learning from any one condition. Both training and distractor images presented in each of the four set sizes were mutually exclusive, and the stimulus presentation order was randomized for each participant. No feedback was provided during the experiment.

A.4.1.4 *Analysis*

As a measure of accuracy, the performance of each subject was assessed by calculating the Matthews correlation coefficient (MCC) (Boughorbel et al., 2017; Chicco and Jurman, 2020). MCC has the virtue of incorporating all information in a confusion matrix (i.e. hits, misses, correct rejects and false alarms) similar to d′ and also handling imbalanced class sizes (that d′ does not handle well). Importantly, MCC is also well-suited for dealing with extreme values. d′ is undefined for perfect performance (hence the need for some ad-hoc corrections; Hautus, 1995), whereas MCC produces a meaningful number (1.0) for perfect performance. The main statistical analysis consisted of a 3-way mixed ANOVA with MCC scores as the dependent variable and image type and set size as within-subject factors and age group as between-subjects factors. The data, study materials, and analyses for this study will be made available via a shared location.

A.4.2 *Results*

A.4.2.1 *Visual memory capacity is dependent on image type, set cardinality, and age group*

Figure A.3 depicts visual memory capacity as a function of image type, set cardinality, and age group. Table A.1 shows the corresponding results of a 3-way mixed ANOVA on MCC scores. The ANOVA revealed a significant main effect of image type (i.e., real vs. abstract images) on recognition performance with a large effect size ($p < 0.001$, $\eta_p^2 = 0.907$; see Table A.1), showing markedly higher recognition of 'real' compared to 'abstract' images (see Figures A.3A&B). Thus, meaningful image content facilitates the recognition of complex visual information.

We also found a significant main effect of set cardinality on performance ($p < 0.001$, $\eta_p^2 = 0.797$; Table A.1), with Bonferroni-corrected post-hoc tests revealing that performance differed significantly between each successive set cardinality (see Supplemental Table A.2). Interestingly, we also see that the difference between 'real' and 'abstract' recognition performance grows as the number of images to be recognized increases (interaction between set cardinality and image type: $p < 0.001$, $\eta_p^2 = 0.608$; Figure A.3E; Table A.1), pointing to the possibility of an increased facilitatory effect of semantically meaningful information as set size increases. Given that participants' performance on 'real' images appears largely resilient to set-size (at least up to the maximum set size we used) whereas the recognition of 'abstract' visual information is strongly modulated by it (Figures A.3A&B), the presence of interpretable content may provide the visual system a 'semantic advantage' for dealing with increasing visual memory load.

Table A.1: Tests of within-subject effects and between-subject effects of 3-way mixed ANOVA, carried out in SPSS, reporting type III sum of squares, degrees of freedom, mean squares, F value, p-value and partial eta squared, each based on the Greenhouse-Geisser results as Sphericity was not assumed.

Variable	SS	df	MS	F	p	η_p^2
Within-subject effects:						
cardinality	4.085	2.260	1.808	148.795	< 0.001	0.797
image type	6.907	1.000	6.907	368.973	< 0.001	0.907
cardinality × age group	0.127	2.260	0.056	4.622	0.010	0.108
cardinality × image type	1.208	2.699	0.448	58.990	< 0.001	0.608
image type × age group	0.086	1.000	0.086	4.608	0.038	0.108
cardinality × image type × age group	0.021	2.699	0.008	1.008	0.387	0.026
Between-subject effects:						
age group	0.208	1	0.208	4.607	0.038	0.108

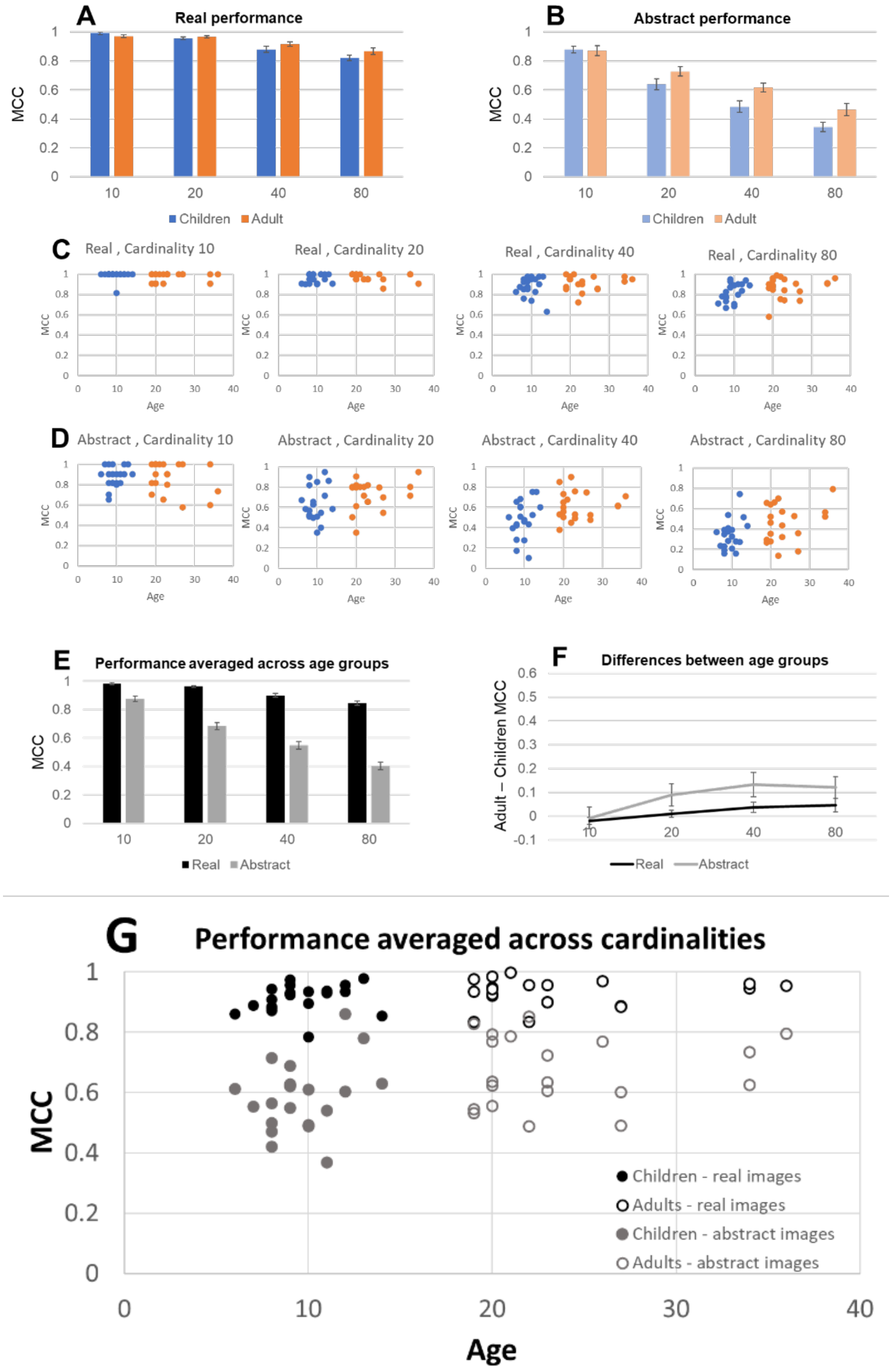


Figure A.3: Visual memory capacity, represented as average MCC values for (A) Real and (B) Abstract images, comparing memory capacity on different image set sizes (10, 20, 40 and 80 images), for both children and adults. (C)&(D) Scatter plots depicting the relationship between age and recognition performance, separately for all combinations of cardinalities (10, 20, 40, and 80) and image type (real, panel (C) and abstract, panel (D)). (E) Contrasting average MCC values on real and abstract images as a function of set size (cardinality), with data pooled across participants from both age groups. (F) Difference in performance between adults and children shown for both real and abstract images. Error bars, in all subpanels (A), (B), (E) and (F), depict standard errors of the mean. (G) Scatter plots depicting the relationship between age and recognition performance averaged across all cardinalities.

The ANOVA also revealed a significant main effect of age group (children vs. adults) on overall recognition performance ($p = 0.038$, $\eta_p^2 = 0.108$; Table A.1 and Figure A.3) as well as a significant interaction effect of age group and set cardinality ($p = 0.010$, $\eta_p^2 = 0.108$; Table A.1) and of age group and image type ($p = 0.038$, $\eta_p^2 = 0.108$; Table A.1). While the former may be explained by the high performance scores achieved with lower set cardinalities masking the differences between age groups that are only reliably observable with higher set cardinalities (Figure A.3D), the latter suggests that robust visual memory capacity may develop earlier for meaningful ('real') as compared to 'abstract' information.

Finer-grained inspection of the experimental data further indicates that in addition to performance differences between groups (i.e., children versus adults), as previously revealed using our 3-way ANOVA (see Table A.1), also within the groups (specifically, within the subgroup of children), higher age appears to be associated with higher performance scores (Figure A.3C&D). The correlation values for the two image classes and all four cardinalities are as follows: $R(10)$ ($r = -0.1515$, $p = 0.3507$), $R(20)$ ($r = 0.0521$, $p = 0.7494$), $R(40)$ ($r = 0.2019$, $p = 0.2116$), $R(80)$ ($r = 0.3282$, $p = 0.0387$), $A(10)$ ($r = -0.1196$, $p = 0.4623$), $A(20)$ ($r = 0.3341, p = 0.0351$), $A(40)$ ($r = 0.3983$, $p = 0.0109$), $A(80)$ ($r = 0.4372$, $p = 0.0048$). Thus, instead of asymptoting within the first few years of life, visual memory capacity appears to keep developing well into late childhood. However, considering the rather small sample size within the children population, this observation will need to be tested more thoroughly in future studies.

A.4.2.2 *Difference in real and abstract performance is not attributable to low level image properties*

We probed whether the observed difference in recognition performance on real vs. abstract images can be explained by differences in low-level image properties or inherent image discriminability across the two image classes. First, to assess the effect of low-level discriminability, we compared the 2D correlation coefficients of all pairwise comparisons across 100 randomly selected real versus 100 randomly selected abstract images (Figure A.3). The two sets were not statistically different ($t(9898) = -0.9364$, $p = 0.3491$, 95% CI = $[-0.0095, 0.0033]$), arguing against differences in low-level discriminability as an account for the observed recognition performance differences.

Next, beyond overall similarity, we extracted specific measures of luminance, spatial frequency, and chromatic content, and compared their variance distributions between the two image classes. With this analysis, we set out to examine if real images may be more discriminable from each other (due to its distribution exhibiting greater variance), relative to abstract images, along any of the outlined image-level dimensions. As Figure A.4 shows, in the case of luminance, we found that the distributions are remarkably similar. To quantify this, we ran Levene's test of the equality of variances with 10000 random subsets of 10, 20, 40, and 80 images per image type (abstract vs. real) each. This analysis revealed that only 5.8%, 5.6%, 5.3%, and 5.4% (for set cardinalities of 10, 20, 40, and 80, respectively) of the 10000 subsets significantly ($p < 0.05$) differed in their variance. With regard to spatial frequency and chromatic content (see Figures A.4D&E), 13.1%, 20.5%, 36.1%, and 63.0% (for set cardinalities of 10, 20, 40, and 80, respectively) as well as 15.7%, 24.6%, 40.4%, and 65.0% (for set cardinalities of 10, 20, 40, and 80, respectively) had significantly different variances. However, while the spatial frequency and chromatic content distributions thus differ in terms of their variances, we note that this observed effect actually renders our behaviorally-observed recognition results more unexpected as the variance in the abstract set, in fact, exceeds that in the real set for all of the three dimensions examined, rather than the other way around.

Taken together, the differences in visual memory capacity of real vs. abstract images are unlikely to be accounted for by differences in the low-level image properties of either stimulus set, and instead are more likely to be attributable to the facilitatory effect of semantics on visual memory.

A.5 COMPUTATIONAL STUDIES

To probe whether the behaviorally observed age-related differences in humans' recognition performance between real and abstract images may have emerged directly through experience with meaningful/naturalistic imagery, we undertook simulations using a computational model system.

A.5.1 *Methods*

In the computational simulations reported in this paper, we utilized a convolutional neural network (CNN), which, although not a perfect model of the biological system, can serve as a rough proxy for sensory processing and its development. The rationale in the context of this study is that we can expose such a system to naturalistic imagery and train it on recognizing real-world objects, while examining the temporal evolution of certain properties of the network that have relevance for performance on an old-new recognition task.

In a classic convolutional neural network, input images are propagated through a hierarchy of convolutional layers that perform spatial filtering operations and extract progressively more complex features of its input. Through subsequent fully-connected layers, these features are eventually transformed into a classification decision – with the

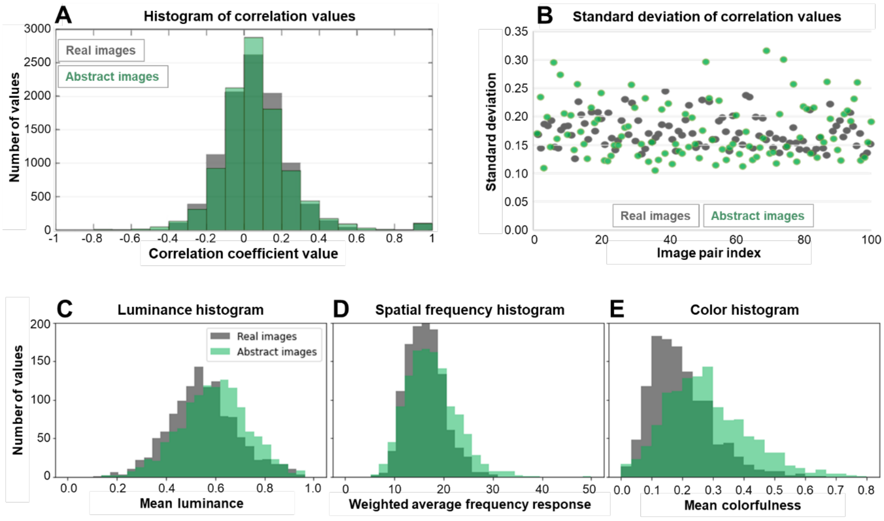


Figure A.4: (A) Histogram of 2D correlation coefficients for all pair-wise comparisons across 100 randomly selected real images (gray distribution) as well as 100 randomly selected abstract images (green distribution). (B) Scatter plot depicting standard deviations of the 2D correlation coefficients, for real as well as abstract images. (C) Histogram of mean luminance scores (computed on the individual pixel level and subsequently averaged across all pixels) for all images in both datasets. (D) Histogram of spatial frequency scores (a 2D Fast Fourier Transform was thereby run on each image after conversion to grayscale, the resulting 2D spectra were radially averaged, and a weighted average of the responses' amplitudes with the corresponding frequency value was computed) for all images in both datasets. (E) Histogram of mean colorfulness scores (computed on the pixel level as imbalance between the R, G, and B channel, and subsequently averaged across all pixels) for all images in both datasets.

final layer of the network, the classification layer, coding for the probabilities that a given input image belongs to each of the categories that the network was trained on differentiating. While, initially, the filters in the convolutional layers and the connection patterns in the fully-connected and final classification layers are random, they get progressively refined during training. As such, one can examine the inner workings of the network, as well as the resulting classification behavior, throughout the entire training process.

To probe whether the empirically observed differences in recognition performance between real and abstract images could be the consequence of experience with real or naturalistic imagery, we trained a typical CNN (specifically, the AlexNet; Krizhevsky et al., 2012) (Figure A.5A) on images belonging to real-world object categories (specifically, a subset of the Caltech-101 image database; Fei-Fei et al., 2004) and, throughout training, examined the network's activations. Specifically, we quantified how different the neural units' activations were when the network was exposed to various sets of naturalistic or abstract test images. The rationale here is that a larger distance in such 'activation space' would render different images in a given set more discriminable from each other, which, in turn, would lead to higher performance in an old-new task, such as the one employed in the human experiments reported above. In other words, we are not explicitly modeling the process of memory in this study but, instead, the differentiation of representations during perceptual analysis – a prerequisite for high recognition performance in an old-new memory task – as a function of the amount of training, which serves as a proxy for developmental change. This rationale is backed up by empirical research showing that humans are better at remembering images that are more distinct from each other (Lukavský and Děchtěrenko, 2017).

The differentiation was thereby quantified as the k-nearest neighbor distance (with k set to 1 for the results reported below, but qualitatively similar results were obtained with other k's) between the activations for a given image and the activations for all other images in the test set. To examine the effects of progressive training on such discriminability, and, thus, its consequences for recognition performance, the analysis was carried out prior to training and after each of the 80 epochs used for training the network. It was run separately for the last four layers in the network (i.e., the last convolutional layer and the three fully-connected layers; Figure A.5A) to also examine potential differences across successive layers of processing. Crucially, to examine differences in inherent discriminability for real vs. abstract images, while considering also the specific role of training, the analysis was carried out for a total of four different test sets. These comprised (i) 50 images of abstract art ("abstract") as well as three different sets of 50 naturalistic, real-world images, belonging to either (ii) 50 real categories from the Caltech-101 database that the network was not

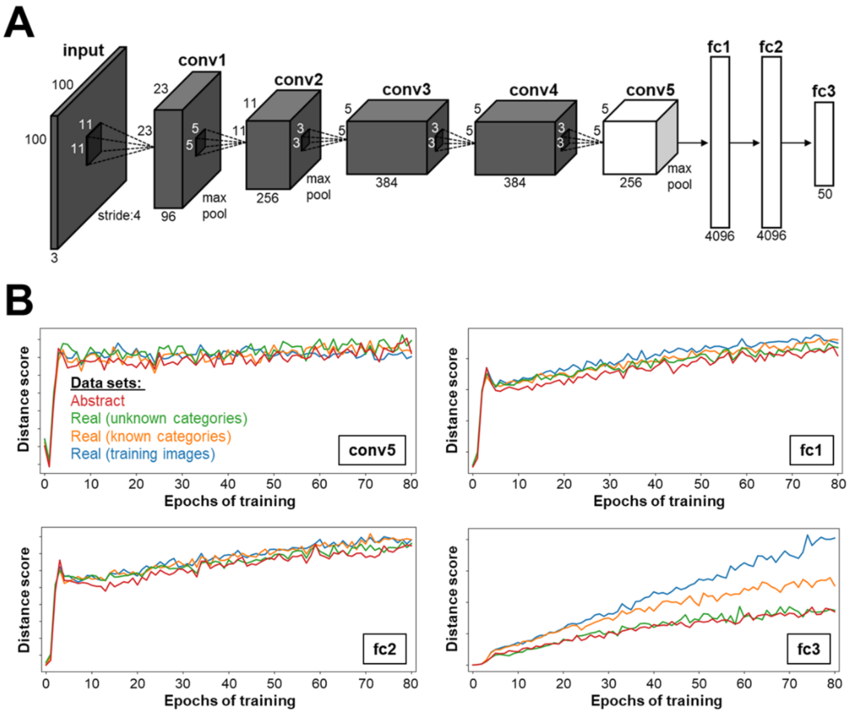


Figure A.5: (A) Architectural sketch of the convolutional neural network utilized for all simulations reported here, with white blocks highlighting the four layers examined for the distance-based analysis shown in B. Note that image sizes were cropped/rescaled to 100x100 pixels and that the input layer dimensions were adjusted accordingly. (B) Distance scores (describing how discriminable the network's activations are for each of the 50 exemplar images sent through it) as a function of training epochs (0 to 80, in steps of 1), network layers (including the last convolutional layer and all three fully-connected layers) as well as the four different image datasets used (50 images of abstract images, 50 images of real images belonging to categories not previously trained on, 50 images of real images belonging to categories previously trained on, and 50 images of real images previously used during training).

trained on ("real (unknown categories)”), (iii) the 50 real categories from the Caltech-101 database that the network was trained on, but not using the exact same images shown during training ("real (known categories)”), and (iv) using 50 of the exact images that the network was previously trained on (“real (training images)”). For sets (ii) to (iv), 50 images were chosen to include one image per category.

Note that to ensure that a greater distance score cannot simply be attributed to larger absolute values that may have naturally been reinforced during training, the activity magnitudes of all except the last fully-connected layer were normalized prior to the analysis (the last fully-connected layer was not normalized as the application of the softmax function already led to an implicit normalization). Additionally, to facilitate the interpretability of differences across image types and as a function of training, the same normalization was applied across all epochs and image sets (but not across layers).

A.5.2 *Results*

We observe that distances generally depict an increase as a function of training epochs (Figure A.5B). This supports the idea that training-induced elaboration of internal representations renders different images in a set more discriminable from each other, predicted to result in improved performance in an old-new task. The timeline and rate of this improvement, however, depends on the layer of processing as well as the image set: whereas the increase of distance scores in the last convolutional layer plateaus after just a few epochs, a steadier improvement can be observed in the fully-connected layers, with the final fully-connected layer (i.e., the one used for classification) exhibiting the highest rate of increase.

Examining relative differences in these curves as a function of image set reveals two key findings. First, for all except the last fully-connected layer, the four image sets yield similar distance score trajectories, suggesting that experience with naturalistic, real-world imagery induces the general shaping of representations in the hidden layers that renders different images in a set more discriminable from each other, regardless of the specific style of the presented imagery (i.e., whether the images are abstract or real) or the similarity of these images to the training set (i.e., whether the network has previously been exposed to the same exemplars, to other exemplars of the same categories, or to neither). Second, in the last fully-connected layer, which is utilized by the network for classification, clear differences between the image sets emerge as a function of training epochs. These differences, however, primarily depend on the similarity of the image sets to the training data rather than on the image style per se: the strongest increase in distance scores as a function of training epochs is evident for the specific set of real images used during training, the second-strongest increase for real images that were not used during training but belong to the trained-on categories, and the lowest increase is seen for the set of real images belonging to unknown categories as well as the set of abstract images. These results allow for interesting inferences: When disregarding the role of familiarity or previous exposure, abstract and real images belonging to categories not explicitly trained on are associated with similar levels of activation differentiation in our networks and, thus, similar predicted recognition performances. Akin to the analyses of the two stimulus sets presented in Figure A.4, the computational results, therefore, further add to the idea that, on the basis of individual images and image properties, there are no fundamental differences between the two image sets that would render one set significantly more inherently discriminable than the other. However, when taking into account the role of familiarity, and when comparing the curve representing different images of trained-on naturalistic categories with that of the abstract image set, we do see marked differences. This finding potentially has some bearing on the nature of the behavioral patterns that we observed in human participants' data. Humans, too, are likely to have previously been exposed to, and have semantic descriptors for, the specific types of image categories presented in the natural-image sets during the experiment. Furthermore, meaningfulness is reliant on familiarity. Given these two premises, if one were to take the curve representing different images of trained-on naturalistic categories as a first-order proxy for the naturalistic condition in the human experiments, these results could be offered as partial account for the empirically-observed differences between recognition performance on real vs. abstract images. By extension, considering that the rate of increase in the representational distances as a function of training epochs is higher for real images of trained-on classes than it is for abstract images, these data add plausibility to the idea that it takes more visual experience to reach the same inter-image discriminability for abstract images, relative to real ones. Although training epochs of the computational model system cannot be linked directly to developmental age of humans, the qualitative performance pattern over time exhibited by the computational model is consistent with the human data reported above. This motivates interesting follow-up experiments on the human front, to examine memory performance on natural objects that participants are unfamiliar with.

While the basic computational results presented above may come as no surprise given that the last fully-connected layer is utilized for final classification, and considering that the two image sets associated with the highest increase in distance scores belong to the categories that the network has been explicitly trained to differentiate, it is worth noting that a noticeable increase in distance scores with progressive training can, in the final layer, also be observed for the other two data sets: real images belonging to unknown categories and abstract art images, even though the network has not been trained on differentiating either. To provide a possible explanation for this finding, we examined the 'confidence' of the network in classifying the different test sets (confidence quantified here as the maximum value observed in the final fully-connected layer after the application of the softmax function). Depicting the distribution of confidence scores for each of the four image sets as a function of training epochs, we see a pattern closely resembling that of Figure A.5B: while the two sets of real images belonging to trained-on categories exhibit the highest classification confidence (with the specific trained-on exemplars exhibiting even higher confidence scores), the two other image sets (i.e., real images belonging to untrained categories as well as abstract images) show a less steep, but still marked, increase in classification confidence throughout training (Figure A.6A). This is particularly noteworthy as for abstract images, for which the object classification task should make little sense, as well as for real images belonging to unseen categories, which do not reasonably map onto the categories available during training, the network's classification confidence is quite high. This is further illustrated in Figure A.6B. To sum, while the similar, unexpectedly high confidence scores for abstract images and real images mapping to unknown categories may account for the computational results observed in the final classification layer, they may also point to an interesting divergence from the human data. Whereas humans' lower memory recognition performance on abstract image sets, as compared to recognition performance on real image sets, could be hypothesized to be accounted for by their inability to compress an abstract image into a memorable label, or a collection of such labels, the network – making confident predictions even for abstract images (e.g., classifying a golden painting as a leopard; see Figure A.6B for illustration) – would not suffer from this problem. In fact, a memory system being linked to the perceptual interpretation of a network might, at least for a modest set size, even profit from such confident predictions. It is notable that the computational simulations with unfamiliar real-world categories yield results consistent with Starr et al. (2020) who observed lower performance on un-nameable objects relative to nameable ones.

A.6 DISCUSSION

The primary goal of this study was to examine whether semantic cues differ in their significance for children on the one hand and adults on the other. Our human experiments reveal that both children and adults benefit from meaningfulness, showing higher memory performance for meaningful images compared to abstract patterns with comparable low-level image statistics. This is in agreement with the results outlined in Boucher et al. (2016) where children's memory

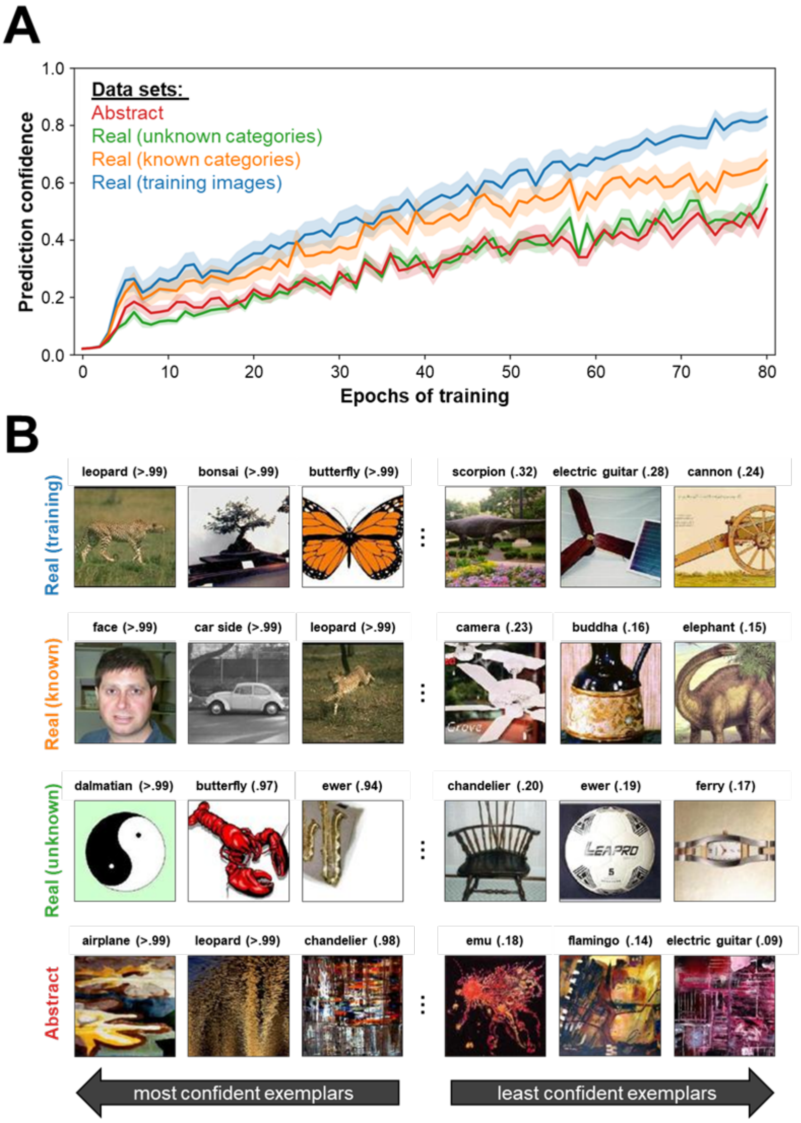


Figure A.6: Mean and standard errors of the confidence in classification prediction (quantified as the maximum value in the final fully-connected classification layer after the application of the softmax function) depicted as a function of training epoch, for each of the four image sets tested. (B) Depiction of the three images associated with the most confident prediction, and the three images associated with the least confident prediction, along with the confidence of the prediction and its label, for each of the three test sets used.

performance was better for concrete than for abstract pictures, marked by larger P600 repetition effects elicited by concrete as compared to abstract designs and a more pronounced and localized recruitment of the left frontal region for concrete pictures. This also agrees with results highlighted in Shoval et al. (2023) where they established that meaning is important for remembering massive amounts of visual information in adults and in Goujon et al. (2022) where they found that semantic information does enhance the encoding and maintaining of images in long-term memory when assessed immediately (but not when assessed after weeks). Building on these findings, our comparative data on children and adults' performance additionally shows that the 'concrete-superiority effect' or the relative advantage conferred by meaningfulness is higher for children. Specifically, while children and adults show comparable recognition performance on meaningful images, the former are significantly worse than the latter when presented with abstract images. In other words, the reduction of semantic cues proves more detrimental for children relative to adults.

The near equality of performance of the two groups with meaningful images, even in the highest cardinality condition, suggests that differences in memory capacity per se are unlikely to provide a satisfactory account of the data with abstract images. Results from our computational work provide a possible explanation for this human finding. If the representational distances between the images in the abstract class are smaller i.e. images of the abstract class are more alike (more confusable with each other) than the meaningful ones, then the greater confusability of the abstract images could manifest as lower accuracy during the recognition phase.

Our computational results further support our human findings. We find that training the computational system on real/meaningful categories later facilitates their discrimination from distractors, thus yielding better performance in an old-new task. Interestingly, when tracking this change as a function of training epochs, which we use as a rough proxy for visual experience during development, we find that in the later stages of the network, the rate of increase in representational distance is higher for real images drawn from the classes the system has been exposed to compared to abstract images. Thus, to obtain the same level of discriminability between images, the timeline is more protracted for abstract images than for familiar, real ones. Translating this to the developmental arena, one would predict that children will take longer to catch up with adults on the set of abstract images relative to real images, consistent with what we observe in our participant data. It is important to note that these differential rates of increase in representational distance are not evident in the early network layers, suggesting that both classes of images are equally well represented by the initial featural vocabulary, as instantiated in the convolutional layers; the differences become apparent in higher-order layers which likely encode more class-specific information.

Some important caveats need to be noted. Firstly, while we have considered the distinction between abstract and real images as being based in the semantic meaninglessness versus meaningfulness of the two sets of images, there is a related, but subtly different, interpretation rooted in the construct of familiarity. Real-world images, by definition, display entities that observers are more familiar with than the abstract patterns. Hence, the differences in performance we observe between real and abstract images could potentially arise from the differing levels of familiarity with the entities depicted in each set. Familiarity may also play a role in modulating performance as a function of age, in that people acquire more familiarity with real-world images as they age. It is also possible that meaningful images, which are more nameable, may be dual-coded (as visual and verbal), whereas meaningless images may be encoded only visually as discussed in Goujon et al. (2022) and this strategy may vary between children and adults.

Secondly, the abstract images we used may not circumvent all involvement with past learning. Much like Ebbinghaus' non-sense syllables (Ebbinghaus, 1913), the abstract images used in this study were non-figurative, with no direct connection to any real-world object, but they could still have indirectly-mediated associations (an abstract pattern that reminds a viewer of a familiar landscape, for instance). It is possible that these indirectly-mediated associations were used as an encoding strategy by some participants, and the tendency to do so might have differed as a function of age. Interestingly, Goujon et al. (2022) found that when asked to assign a name to real or abstract images at two time points, people were highly consistent across time with the assignments for real images, but much less so for the abstract patterns. This reduction of consistency would weaken the influence of a naming strategy on the kind of task we have employed here, but would not eliminate it. Hence, this is an interesting avenue for future research. Another caveat relates to the many factors that can contribute to memory performance. These include attention (Shipstead et al., 2015), rehearsal (Kellas et al., 1975; Rundus, 1971), association (Voss, 2009), methods of learning (Meumann, 1913; Qureshi et al., 2014), and fatigue (Finkenbinder, 1913). The developmental changes in performance that we have observed may, therefore, arise not only from the representational elaboration that we referred to above, but also possibly from changes in any of these factors. However, this likelihood is mitigated by the statistically indistinguishable performance of the pediatric and adult groups on the set of natural images suggesting that basic skills such as task understanding and attentiveness are comparable across the two groups.

This work points to several potentially interesting follow-on studies. These include: 1. Examining whether memory performance with meaningful images showing unfamiliar objects/settings is closer to that corresponding to abstract images (as is the case for computational simulations) or the real images, 2. Studying how larger time delays between memorization and recognition impact recognition performance. Are the compact linguistic descriptions associated with meaningful images more resilient to decay over time than the eidetic representations that abstract images require? Specifically, does the passage of time lead to a greater decay of recognition performance for one type of images, such as the abstract set, than for the other? 3. Investigating whether the intra- and inter-group results presented here would differ had participants been tasked with memorizing more than 80 pictures. 4. Working with younger children to help reveal when, in the developmental timeline, differences in performance between real and abstract images first become evident.

A.7 CONCLUSION

How meaning influences memory has been an active and fruitful area of research both in the domains of word recall and vision and it is known that meaning does have a facilitatory effect on memory for both short and long durations. It is however not known how this 'semantic benefit' varies as a function of age. We looked at how long-term visual memory capacity is modulated by semantics and how it changes as a function of age in children and young adults and also performed computational experiments to complement our human studies. Results of human studies show a clear benefit of semantics on the memory performance of both children and adults and an age-dependent improvement in performance, especially for abstract images. Our computational results help partly account for the human data but also point to, and make predictions for, new experiments.

ACKNOWLEDGEMENTS

The authors wish to thank all the participants and their families who contributed to this study.

A.8 SUPPLEMENTAL MATERIAL

REFERENCES

Asp, Isabel E, Viola S Störmer, and Timothy F Brady (2021). "Greater visual working memory capacity for visually matched stimuli when they are perceived as meaningful." In: *Journal of cognitive neuroscience* 33.5, pp. 902–918.

Table A.2: (Supplemental Table) Bonferroni-corrected pairwise dependent sample t-tests related to the main effect of set cardinality.

Cardinality	Cardinality	**P-value**	95% CI Lower bound (Bonferroni-adjusted)	95% CI Upper bound (Bonferroni-adjusted)
10	20	**<0.001**	0.065	0.146
10	40	**<0.001**	0.164	0.244
10	80	**<0.001**	0.251	0.357
20	40	**<0.001**	0.058	0.139
20	80	**<0.001**	0.146	0.251
40	80	**<0.001**	0.067	0.132

Baddeley, Alan D and Graham J Hitch (2017). "Is the levels of processing effect language-limited?" In: *Journal of Memory and Language* 92, pp. 1–13.

Bellhouse-King, Mathew W and Lionel G Standing (2007). "Recognition memory for concrete, regular abstract, and diverse abstract pictures." In: *Perceptual and motor skills* 104.3, pp. 758–762.

Boucher, Olivier, Christine Chouinard-Leclaire, Gina Muckle, Alissa Westerlund, Matthew J Burden, Sandra W Jacobson, and Joseph L Jacobson (2016). "An ERP study of recognition memory for concrete and abstract pictures in school-aged children." In: *International Journal of Psychophysiology* 106, pp. 106–114.

Boughorbel, Sabri, Fethi Jarray, and Mohammed El-Anbari (2017). "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric." In: *PloS one* 12.6, e0177678.

Bower, Gordon H, Alan M Lesgold, and David Tieman (1969). "Grouping operations in free recall." In: *Journal of Verbal Learning and Verbal Behavior* 8.4, pp. 481–493.

Brady, Timothy F and Viola S Störmer (2022). "The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 48.7, p. 942.

Brainard, David H and Spatial Vision (1997). "The psychophysics toolbox." In: *Spatial vision* 10.4, pp. 433–436.

Calkins, Mary Whiton (1898). "Short studies in memory and in association from the Wellesly College Psychological Laboratory." In: *Psychological Review* 5.5, p. 451.

Chicco, Davide and Giuseppe Jurman (2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." In: *BMC genomics* 21.1, pp. 1–13.

Collins, Allan M and M Ross Quillian (1969). "Retrieval time from semantic memory." In: *Journal of verbal learning and verbal behavior* 8.2, pp. 240–247.

Cowan, Nelson, Angela M AuBuchon, Amanda L Gilchrist, Timothy J Ricker, and J Scott Saults (2011). "Age differences in visual working memory capacity: Not based on encoding limitations." In: *Developmental science* 14.5, pp. 1066–1074.

Dooling, D James and Roy Lachman (1971). "Effects of comprehension on retention of prose." In: *Journal of experimental psychology* 88.2, p. 216.

Ebbinghaus, Hermann (1913). *Memory: A Contribution to Experimental Psychology; Translated by Henry A. Ruger and Clara E. Bussenius*. Teachers College, Columbia University, New York.

Ernest, Carole H and Allan Paivio (1971). "Imagery and verbal associative latencies as a function of imagery ability." In: *Canadian Journal of Psychology/Revue canadienne de psychologie* 25.1, p. 83.

Fei-Fei, Li, Rob Fergus, and Pietro Perona (2004). “Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories.” In: *2004 conference on computer vision and pattern recognition workshop*. IEEE, pp. 178–178.

Ferrara, Katrina, Sarah Furlong, Soojin Park, and Barbara Landau (2017). “Detailed visual memory capacity is present early in childhood.” In: *Open Mind* 2.1, pp. 14–25.

Finkenbinder, Erwin Oliver (1913). “The curve of forgetting.” In: *The American Journal of Psychology* 24.1, pp. 8–32.

Forsberg, Alicia, Eryn J. Adams, and Nelson Cowan (2022). “The development of visual memory.” In: *Visual Memory*. Routledge. Chap. chapter17.

Goldstein, Alvin G and June E Chance (1971). “Visual recognition memory for complex configurations.” In: *Perception & Psychophysics* 9, pp. 237–241.

Goujon, Annabelle, Fabien Mathy, and Simon Thorpe (2022). “The fate of visual long term memories for images across weeks in adults and children.” In: *Scientific Reports* 12.1, p. 21763.

Hautus, Michael J (1995). “Corrections for extreme proportions and their biasing effects on estimated values of d′.” In: *Behavior Research Methods, Instruments, & Computers* 27, pp. 46–51.

Kellas, George, Charley McCauley, and Carl E McFarland Jr (1975). “Development aspects of storage and retrieval.” In: *Journal of Experimental Child Psychology* 19.1, pp. 51–62.

Konkle, Talia, Timothy F Brady, George A Alvarez, and Aude Oliva (2010). “Conceptual distinctiveness supports detailed visual long-term memory for real-world objects.” In: *Journal of experimental Psychology: general* 139.3, p. 558.

Koutstaal, Wilma, Chandan Reddy, Eric M Jackson, Steve Prince, Daniel L Cendan, and Daniel L Schacter (2003). “False recognition of abstract versus common objects in older and younger adults: testing the semantic categorization account.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.4, p. 499.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). “Imagenet classification with deep convolutional neural networks.” In: *Advances in neural information processing systems* 25.

Lukavskỳ, Jiří and Filip Děchtěrenko (2017). “Visual properties and memorising scenes: Effects of image-space sparseness and uniformity.” In: *Attention, Perception, & Psychophysics* 79, pp. 2044–2054.

Madigan, Stephen (2014). “Picture memory.” In: *Imagery, memory and cognition*, pp. 65–89.

Meumann, Ernst (1913). *The psychology of learning: An experimental investigation of the economy and technique of memory*. Appleton.

Moulin, Christopher JA, Martin A Conway, Rebecca G Thompson, Niamh James, and Roy W Jones (2005). “Disordered memory awareness: recollective confabulation in two cases of persistent déjà vecu." In: *Neuropsychologia* 43.9, pp. 1362–1378.

Olson, GM (1976). *An information processing analysis of visual memory and habituation in infants. InT. J. Tighe & RN Leaton (Eds.), Habituation: perspectives from child development, animal behavior, and neurophysiology. Hills-dale*.

Paivio, Allan (1971). "Imagery and language." In: *Imagery*. Elsevier, pp. 7–32.

Pezdek, Kathy (1977). "Cross-modality semantic integration of sentence and picture memory." In: *Journal of Experimental Psychology: Human Learning and Memory* 3.5, p. 515.

Qureshi, Ayisha, Farwa Rizvi, Anjum Syed, Aqueel Shahid, and Hana Manzoor (2014). "The method of loci as a mnemonic device to facilitate learning in endocrinology leads to improvement in student performance as measured by assessments." In: *Advances in physiology education* 38.2, pp. 140–144.

Riggs, Kevin J, James McTaggart, Andrew Simpson, and Richard PJ Freeman (2006). "Changes in the capacity of visual working memory in 5-to 10-year-olds." In: *Journal of experimental child psychology* 95.1, pp. 18–26.

Rose, Susan A, Judith F Feldman, and Jeffery J Jankowski (2001). "Visual short-term memory in the first year of life: capacity and recency effects." In: *Developmental psychology* 37.4, p. 539.

Ross-sheehy, Shannon, Lisa M Oakes, and Steven J Luck (2003). "The development of visual short-term memory capacity in infants." In: *Child development* 74.6, pp. 1807–1822.

Rundus, Dewey (1971). "Analysis of rehearsal processes in free recall." In: *Journal of experimental psychology* 89.1, p. 63.

Sasson, Ralph Y and Paul Fraisse (1972). "Images in memory for concrete and abstract sentences." In: *Journal of Experimental Psychology* 94.2, p. 149.

Shipstead, Zach, Tyler L Harrison, and Randall W Engle (2015). "Working memory capacity and the scope and control of attention." In: *Attention, Perception, & Psychophysics* 77, pp. 1863–1880.

Shoval, Roy, Nurit Gronau, and Tal Makovski (2023). "Massive visual long-term memory is largely dependent on meaning." In: *Psychonomic Bulletin & Review* 30.2, pp. 666–675.

Simmering, Vanessa R (2012). "The development of visual working memory capacity during early childhood." In: *Journal of experimental child psychology* 111.4, pp. 695–707.

Slamecka, Norman J (1985). "Ebbinghaus: Some associations." In.

Slater, Alan (1989). "Visual memory and perception in early infancy." In: *Infant development*, pp. 43–71.

Standing, Lionel (1973). "Learning 10000 pictures." In: *Quarterly Journal of Experimental Psychology* 25.2, pp. 207–222.

Starr, Ariel, Mahesh Srinivasan, and Silvia A Bunge (2020). "Semantic knowledge influences visual working memory in adults and children." In: *PloS one* 15.11, e0241110.

Tulving, Endel et al. (1972). "Episodic and semantic memory." In: *Organization of memory* 1.381-403, p. 1.

Voss, Joel L (2009). "Long-term associative memory capacity in man." In: *Psychonomic Bulletin & Review* 16, pp. 1076–1081.

Walker, Peter, Graham J Hitch, Alison Doyle, and Tracey Porter (1994). "The development of short-term visual memory in young children." In: *International Journal of Behavioral Development* 17.1, pp. 73–89.

B
RECOGNIZING DISTANT FACES

Content from

Jarudi, I. N., Braun, A., **Vogelsang, M.**, Vogelsang, L., Gilad-Gutnick, S., Bosch, X. B., Dixon, III, W. V., & Sinha, P. (2023). "Recognizing distant faces". **Published** in Vision Research, 205, 108184.

B.1 ABSTRACT

As an 'early alerting' sense, one of the primary tasks for the human visual system is to recognize distant objects. In the specific context of facial identification, this ecologically important task has received surprisingly little attention. Most studies have investigated facial recognition at short, fixed distances. Under these conditions, the photometric and configural information related to the eyes, nose and mouth are typically found to be primary determinants of facial identity. Here we characterize face recognition performance as a function of viewing distance and investigate whether the primacy of the internal features continues to hold across increasing viewing distances. We find that exploring the distance dimension reveals a qualitatively different salience distribution across a face. Observers' recognition performance significantly exceeds that obtained with the internal facial physiognomy, and also exceeds the computed union of performances with internal and external features alone, suggesting that in addition to the mutual configuration of the eyes, nose and mouth, it is the relationships between these features and external head contours that are crucial for recognition. We have also conducted computational studies with convolutional neural networks trained on the task of face recognition to examine whether this representational bias could emerge spontaneously through exposure to faces. The results provide partial support for this possibility while also highlighting important differences between the human and artificial system. These findings have implications for the nature of facial representations useful for a visual system, whether human or machine, for recognition over large and varying distances.

KEYWORDS

Face recognition; Long-range viewing; Internal and external features; Image degradations

B.2 INTRODUCTION

There exists a notable discord between what we know of the spatial aspects of real-world social interactions on the one hand, and the foci of investigation in the domain of face recognition on the other. Over the past 60 years, the field of 'proxemics' has characterized how an individual's personal space is organized. While the early studies (Hall, 1966) were rather qualitative and somewhat weak in empirical rigor, they provided a useful taxonomy of space corresponding to the nature of interactions they allow. For instance, Hall (1966), defines 'far-phase' of public space as a distance where "subtle shades of meaning conveyed by the normal voice are lost as are the details of facial expression and movement." An individual is able to gauge different aspects of people as they move through these zones of space. The critical initial recognition step of this process is believed to occur at approximately 50 feet (Fotios et al., 2018). Even at this extended distance, the identity decision is sub-served, in large part, by facial cues rather than body structure (Burton et al., 1999). The initial step of facial recognition can help facilitate a decision about whether to approach or avoid the person, resulting in either reduced or increased inter-personal distance. Consider, for instance, recognizing a friend in the large arrival hall of an airport, or crossing over to the other side of the road to avoid meeting a garrulous colleague whom you spot sauntering a block away. In spatial terms, the progression from recognition to interaction (assuming an 'approach' decision) translates to transitioning from long ranges in 'public spaces' to potentially up-close 'personal spaces', as shown in Fig. B.1a.

In contrast to the aforementioned ecological importance of face recognition at large distances, laboratory studies have largely focused on examining facial recognition at 'up-close' conditions. The stimuli typically used in these studies comprise large, high-resolution images, quite unlike the information available when viewing distant faces. For instance, seeing a high-resolution 6 degrees wide face image, a stimulus size that many past studies have used or exceeded (e.g., Andrews et al., 2010; Burton et al., 2005; Maurer et al., 2002; O'Donnell and Bruce, 2001; Schwaninger et al., 2003; Yovel and Kanwisher, 2004), is tantamount to viewing a real face from a distance of approximately 4 feet (as per calculations in Oruc et al., 2019). This corresponds to the viewed individual being within the viewer's 'personal space' (Hall, 1966), which is typically intended for interaction with an already recognized person. It is indeed the case that most of the interactions adults have with others are conducted up-close. Oruc et al. (2019) analyzed first-person videos collected with head-mounted cameras and found that the majority of faces experienced had angular widths of 6 degrees or more, with familiar faces subtending, on average, larger angles than unfamiliar ones. Hence, the choice of large face

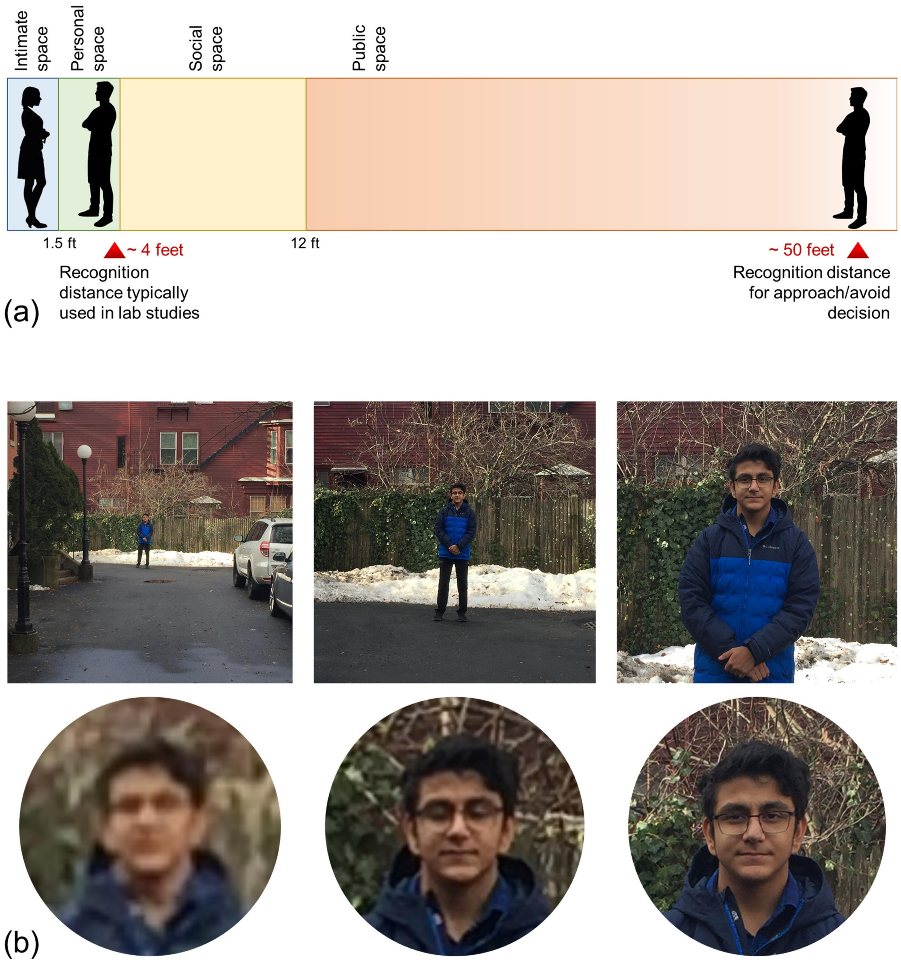

Intimate space
Personal space
Social space
Public space
1.5 ft
~ 4 feet
Recognition distance typically used in lab studies
12 ft
~ 50 feet
Recognition distance for approach/avoid decision
(a)
(b)

Figure B.1: (a) A depiction of the conception of interpersonal distances by Hall (1966). (b) Viewing distance determines the size of the image projected on the retina (upper panels) and thereby changes the effective resolution available (lower row).

stimuli in the aforementioned stimuli appears to be consistent with ecological statistics. However, by focusing on identification within the personal space, laboratory studies run the risk of neglecting settings involving large viewing distances in which human face recognition often transpires. Given that many people transitioning through an individual's public space are likely to be strangers, they will not elicit recognition, but the ability to identify the few who are indeed familiar is necessary to encourage further approach and admittance to the personal space. It is worth noting that this argument of the significance of recognition at a distance applies to identities with whom one is already familiar. Familiarization with new identities is a process that may well rely on face exposures at close distances. While such face-learning is a topic of great interest, here we are concerned exclusively with recognition of already familiar individuals.

It is important to note that we cannot assume that results of conventional studies will necessarily prove to be applicable to understanding face recognition over much greater distances. Distance induces transformations on the facial images projected on the viewer's retinas, notably reductions in resolution. Thus, cues that may be available up-close, may simply not exist at greater distances, forcing the visual system to adopt different recognition strategies in such settings.

While existing literature on face recognition as a function of distance is sparse, several studies have examined the related issue of the roles of different spatial frequencies. Broadly, these investigations have revealed that low spatial frequencies allow for configural processing, while high spatial frequencies support featural analysis by providing information about details (Cheung et al., 2008; Fiorentini et al., 1983; Goffaux et al., 2005; Oliva et al., 2006; Parker et al., 1996; Ruiz-Soler and Beltran, 2006). Further, Mousavi and Oruc (2020) have examined the effect of stimulus size on blurry face recognition and revealed a scale-dependent influence on recognition resilience to blur. Additional studies have directly considered the viewing distance dimension (De Jong et al., 2005; Greene and Fraser, 2002; Hahn et al., 2016; Lampinen et al., 2014; Lindsay et al., 2008; Loftus and Harley, 2005; Wagenaar and Van Der Schrier, 1996). These studies have primarily focused on determining the range over which above-chance recognition can be obtained, rather than examining the contribution of different facial cues to observers' performance. It is this latter goal that we focus on in this paper. To this end, we use as a starting point a prominent result that has consistently emerged across numerous studies of 'up-close' face recognition, and then investigate whether and how this result changes as viewing distance increases, namely, the role of internal facial physiognomy, viz., the featural details and mutual configuration of the eyes, nose and mouth (Burton et al., 2005; Ellis et al., 1979; Hosie et al., 1988; Maurer et al., 2002; O'Donnell and Bruce, 2001; Valentine, 1991; Young et al., 1985; Yovel and Kanwisher, 2004). The importance of these features is widely accepted, particularly in the identification of familiar faces, and is even reflected in popular literature. For example, the main protagonist in Isabelle Holland's eponymous 1928 book is referred to as 'the man without a face' because he suffered a burn injury that disfigured his eyes, nose and mouth.

Building on this observed importance of the internal facial physiognomy for 'up close' faces, a question we are interested in addressing is whether this cue continues to inform recognition across larger viewing distances as well, or if the visual system switches to using other facial cues. As viewing distance increases, the amount of detailed featural information in a face progressively decreases (see Fig. B.1b), rendering it harder to decode identity from the internal features (Loftus and Harley, 2005). However, it is possible that all facial identity cues degrade in a similar manner, and hence the relative importance of the internal features is maintained irrespective of the viewing distance. Alternatively, increasing viewing distance may impact different aspects of facial information differently, leading to a change in the relative importance of facial cues as a function of distance. We seek to directly address this issue through studies of familiar face recognition across different viewing distances.

Our studies have three specific aims. First, we characterize overall face recognition performance as a function of viewing distance. Second, we examine whether overall face recognition performance at any given viewing distance can be largely accounted for by internal facial features. Finally, we undertake computational simulations with deep neural networks to determine whether the empirically observed featural bias in human responses may emerge spontaneously through exposure to a large training set of faces.

The structure of the paper is as follows. We start by describing a study that sought to examine whether a proxy for viewing distance – Gaussian blur – reveals potentially interesting results in terms of the effectiveness of facial cues at different magnitudes of image transformation. This is followed by a study that directly investigates face recognition as a function of viewing distance. To foreshadow the results from these two studies, we find that more than the mutual configuration of the eyes, nose and mouth, it is the relationships between these features and external head contours that are crucial for recognition at a distance. Finally, we describe our computational investigations that probe whether the featural biases exhibited by human subjects can arise spontaneously in a machine vision system through experience with several faces. These results provide partial support for this possibility while also highlighting important differences between the human and artificial system.

B.3 HUMAN EXPERIMENT 1: ROBUSTNESS OF INTERNAL FEATURES FOR FACIAL IDENTIFICATION AS A FUNCTION OF INCREASED BLUR

Increasing viewing distance leads to a diminishment of the image size projected onto the retina, thereby resulting in reduced resolution, given that fewer photoreceptors are stimulated. Loftus and Harley (2005) have detailed this linkage between viewing distance and blur. Working with a set of four synthetically generated faces, they showed that recognition performance at increasing distances was analogous to recognition of increasingly blurred images. However, this linkage is an approximate one since the reduction in image quality that results from actually increasing viewing distance between the observer and observed is not precisely the same as convolving with a Gaussian an image obtained up-close. Several factors lead to this imperfect equivalence. These include ocular micro-saccades that smear image information more for distant objects than those up-close, intervening haze, and discretization artifacts introduced when sampling very small images. Nevertheless, the linkage between distance and blur is useful as a first approximation for determining the effects of the former. Given the relative simplicity of experimental design and logistics, we decided to start by exploring face recognition performance and cue effectiveness as a function of image blur. These results would help set the stage for more directly probing the viewing distance dimension.

B.3.1 *Methods*

Our stimulus set consisted of twenty-four high-resolution color images of famous individuals in frontal views. The celebrities were movie or television actors and politicians. All twenty-four faces used were scale-normalized to have an inter-pupillary distance of 50 pixels. The heads were rotoscoped so that all images had a uniform white background. From this collection, we created a total of four stimulus sets using Adobe Photoshop:

- Set A: Internal features placed in a row, not preserving their mutual spatial configuration.
- Set B: Internal features in their correct spatial configuration.
- Set C: External features alone.
- Set D: The original collection of whole-head images.

Samples of stimuli used for each of the experiments are shown in Fig. B.2. To measure recognition as a function of increasing blur, images in each set were subjected to several levels of Gaussian blur

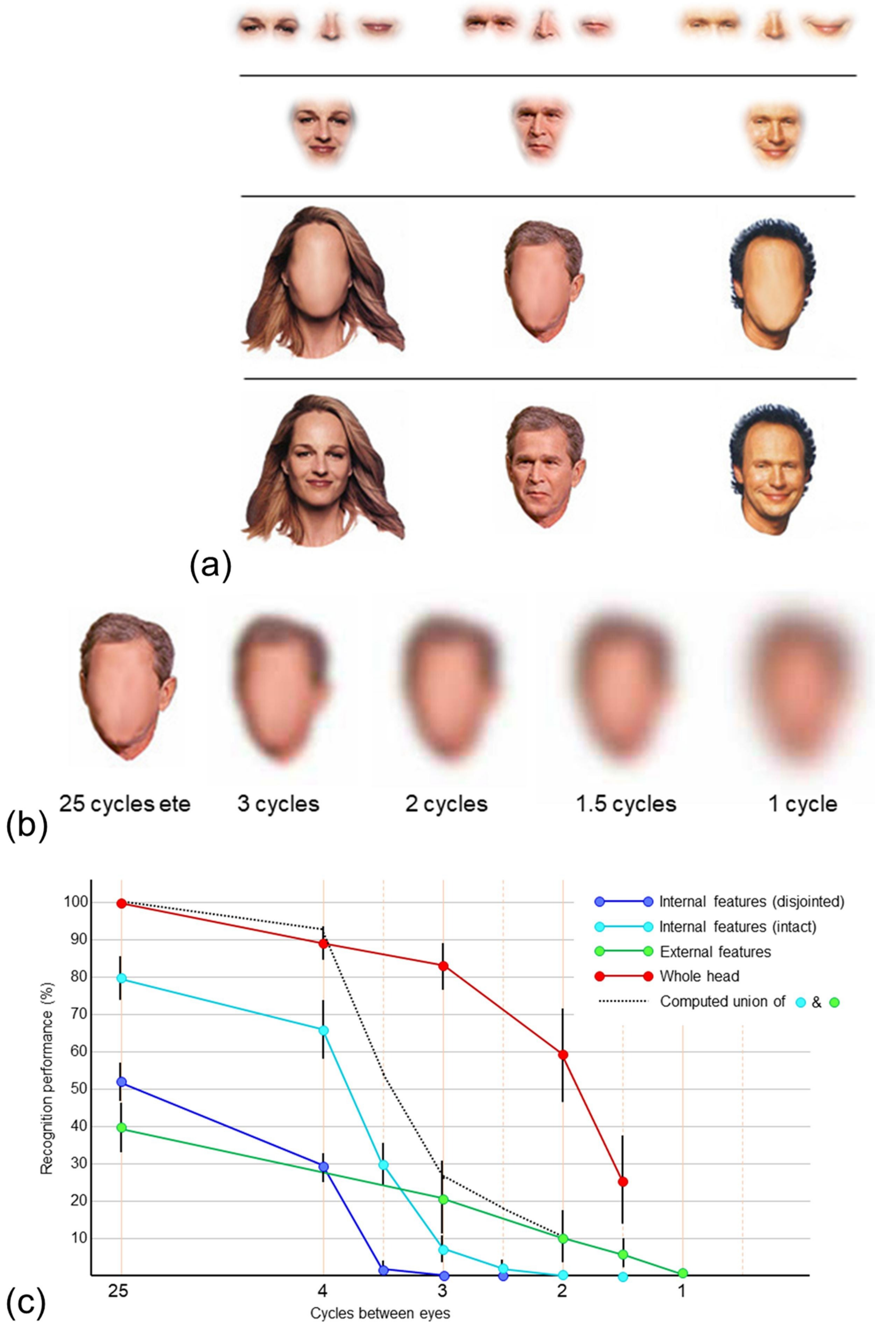


Figure B.2: (a) Sample stimuli from the four conditions, from top to bottom: Condition A: Internal features for each face placed in a row. Condition B: Internal features for each face in their correct spatial configuration. Condition C: External features alone with the internal features digitally erased. Condition D: Full faces, including both internal and external features. (b) Sample stimuli to illustrate the impact of different levels of blur. The annotations under each image indicate the number of cycles eye to eye. (c) Recognition performance as a function of image resolution across the four stimulus conditions. The dashed curve represents the computed upper-bound of the union of performances obtained with the intact internal and external features. The error bars represent 95 % confidence intervals.

with the resulting resolutions ranging from 1 cycle between the eyes to 4 cycles.

Thirty subjects, ranging in age from 18 to 38, participated in the study. This study was approved by Massachusetts Institute of Technology's IRB and participants gave informed consent. Subjects were randomly placed in four non-overlapping groups, corresponding to the four experimental conditions (eight each in experiments A through C, and six in experiment D). The mutual exclusion was enforced to prevent any transfer of information from one condition to another.

In each experimental condition, the presented stimuli proceeded from the most degraded to the least degraded conditions. Subjects viewed the screen from approximately 1.97 feet (60 cm) but were allowed to move their heads freely. At around 2 feet, a face subtended, on average, 5 degrees of visual angle. Subjects were asked to identify each individual shown either by name or some uniquely identifying information (such as a job for which the celebrity may have been famous; for example, for actors this would include the name of a movie, television show or character with which he or she may have been associated). In each experimental condition, un-degraded faces from set D were shown subsequent to the presentation of all other stimuli as a reference set, allowing us to normalize our data at the individual subject level, since recognition of an undegraded full head image indicated that the subject was actually familiar with that particular face under normal conditions. Subsequent data analysis considered recognition data only for those face images that each subject was able to identify in this reference set.

B.3.2 *Results*

Fig. B.2c plots performance (i.e., the proportion of faces correctly identified) as a function of image blur. The graph shows four curves corresponding to each of the experimental conditions tested (A through D).

Performance in the full-face condition (condition D) is remarkably robust to reductions in image quality and declines only slowly with increasing image blur. Even at a blur level resulting in only 3 cycles between the eyes, performance is greater than 80 %. This is in contrast to performance with the rearranged internal features. In this condition (condition A), even at the highest resolution, performance is rather modest, averaging just a little over 50 % and dropping rapidly to essentially 0 % with increasing blur. When the internal features are placed in their correct spatial configuration (condition B), performance improves relative to condition A, but continues to be extremely sensitive to the amount of blur applied to the stimulus images.

Our data also indicate that in the high-resolution case, simply excluding external features severely compromises recognition, consistent with previous findings (Lewin and Herlitz, 2002; Megreya and Bindemann, 2009; Wright and Sladden, 2003). In fact, in contrast to the rapid fall-off observed for internal features only (conditions A&B), the shallow slope of the curve for external features alone (condition C) indicates gradual performance change with increasing blur. However, the absolute level of performance all along this latter curve is poor, not exceeding 40 % even at the highest resolution. Interestingly, at a resolution of approximately 3.5 cycles between the eyes, we observe a change in the rank-ordering of the curves for recognition of internal-only versus external-only features (i.e., condition C relative to conditions A and B). These results are consistent with past work showing that while high spatial frequencies support detailed featural analysis, lower spatial frequencies have been related to more configural processing (Cheung et al., 2008; Fiorentini et al., 1983; Goffaux et al., 2005; Oliva et al., 2006; Parker et al., 1996; Ruiz-Soler and Beltran, 2006). Although the stimuli used in condition D can be obtained by a superposition of the stimuli in conditions B and C, the result of summing up performances with conditions B and C falls significantly short of the plot corresponding to condition D.

A 2-factor repeated measures ANOVA on combined data across four blur levels (1.5, 2, 3, and full resolution) and the 4 experimental conditions showed highly significant main effects of blur level ($F_{3,78} = 420.036, p < .001$) and condition ($F_{3,26} = 128.049, p < .001$) as well as a highly significant blur level by condition interaction ($F_{9,78} = 32.870, p < .001$). At high levels of blur (levels 1.5, 2 and 3), performance on the full-face condition was significantly better than in the external features only condition, which was in turn better than either of the internal features only conditions (all $p < 0.05$ in two-tailed t-tests).

Overall, the data strongly suggest that with increasing blur, internal features on their own are inadequate to account for performance with whole heads. Furthermore, mere addition of performance from external features is insufficient to bridge the gap in performance, obviating the possibility of the two sets of cues serving as independent sources of identity information. Instead, it is the mutual spatial relations between the two sets of features that appear to be necessary to account for the observed recognition performance.

We next sought to determine whether this pattern of results would continue to hold when the independent variable was viewing distance, rather than blur as its proxy.

B.4 HUMAN EXPERIMENT 2: OVERALL RECOGNITION AND ROBUSTNESS OF INTERNAL FEATURES FOR FACIAL IDENTIFICATION AS A FUNCTION OF INCREASED VIEWING DISTANCE

Inheriting the basic experimental design from the study on blur, this experiment explored how distance-conditioned recognition performance with full faces relates to that obtained with the internal and external features independently, i.e., the cue-fusion strategy that the visual system uses for combining information from these two sets of features.

B.4.1 *Methods*

20 subjects from the MIT student community participated in the experiments. This study was approved by Massachusetts Institute of Technology's IRB and participants gave informed consent. All participants underwent acuity testing prior to the start of the study. Only those with 20/20 acuity were enrolled in the subsequent experiment. Forty celebrity faces, distributed across the same four conditions as in the first experiment (see Fig. B.1a), were presented on an LCD screen sequentially to subjects in random order, with no identities repeated across conditions for a given subject. Additionally, we included a fifth condition showing the whole heads vertically inverted.

Images were displayed on a monitor placed on a wheeled trolley that could be moved along a 25- foot-long track marked on the floor, marked in inches (the start distance was 25 feet). The experimenter moved the trolley slowly (1 in./second) with the aid of a tether, progressively closer to the observer. The observer's task was to try to identify the individual shown as soon as possible. When they indicated that they were ready to give a response, the trolley was stopped and their response and the trolly's distance was recorded. The trolley resumed moving closer to the observer if they had produced an incorrect response. For correct responses, the trolley was moved closer for three more feet to determine if the subjects would change their response.

The condition that a given celebrity face appeared in was randomized across subjects. For each subject, performance in each condition was computed after normalizing responses against a reference test conducted after the main experiment (similar to experiment 1). In this baseline, subjects were shown, at around 2 feet (60 cm), high-resolution full head images of each celebrity in turn and asked to recognize them. This allowed us to determine which of these individuals the subject was actually familiar with. Those that were not recognized in the baseline condition were not included in the analyses of that particular subject's data from the experiment.

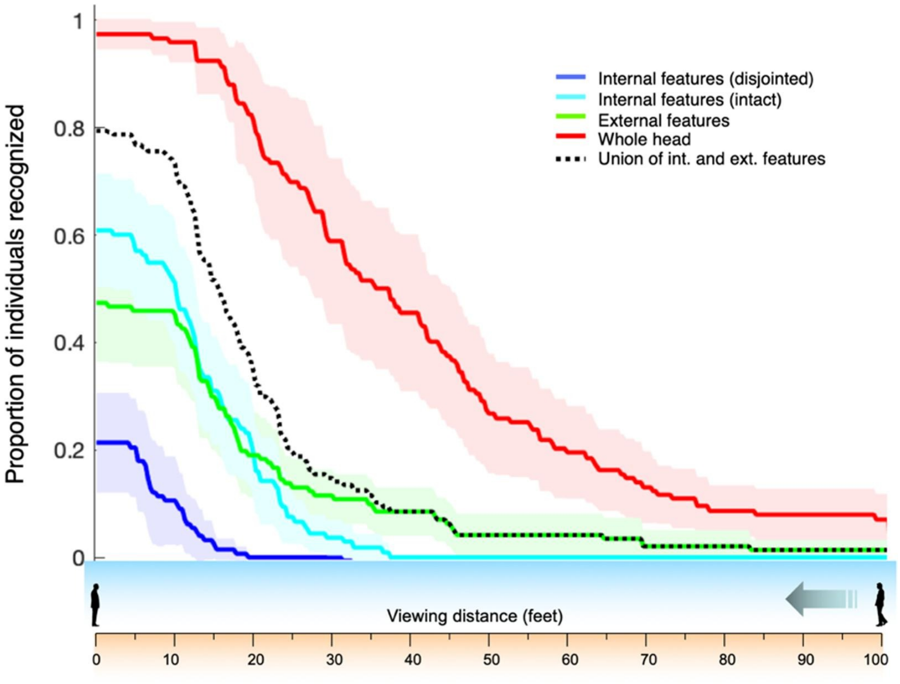


Figure B.3: Recognition performance as a function of image distance across each of the four stimulus conditions. The key finding to note is that the actual performance with full faces (red line) significantly exceeds that predicted by the union of performances obtained with internal and external feature sets independently (dashed black line) at all viewing distances. The shaded regions around each line represent 95 % confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

B.4.2 *Results*

Fig. B.3 shows average performance (i.e., the proportion of faces correctly identified) as a function of viewing distance. The graphs show performance corresponding to each of the experimental conditions tested. Chance performance in all these conditions is close to zero because subjects were unaware of which individuals they might see in the session.

In condition A (rearranged internal features), even at the closest distance (12”), performance is modest, averaging just a little over 20 %, suggesting that on their own, disjointed internal features provide extremely limited information for facial identity. Furthermore, the performance drops sharply with increasing distance. When the internal features are placed in their correct spatial configuration (condition B), performance improves relative to condition A, but continues to be extremely sensitive to the viewing distance. External features, on their own (condition C), also support modest performance (barely exceeding 40 % at the closest distance) and show a marked decline in usefulness beyond 20 feet. In contrast to these findings, the presentation of whole heads (condition D) dramatically improved performance, especially at long viewing distances.

The high performance with internal features at close viewing distances is consistent with past research showing that these features are more useful for familiar face recognition than external features (Ellis et al., 1979; Young et al., 1985) in high-resolution images. Furthermore, in settings of close-up viewing, the union of conditions B and C yields high performance accuracy, as shown in Fig. B.3. On its own, this finding is consistent with the internal and external features making independent contributions to the overall process of identity computation. However, an interesting result emerges as we proceed further along the distance dimension and compare the computed performance (union of B and C) with the actual performance obtained with whole head stimuli. Although the stimuli used in condition D can be obtained by a superposition of the stimuli in conditions B and C, the result of summing up the performances of conditions B and C (calculated as the union) still falls significantly short of the actual performance obtained in condition D (one-tailed two-sample KS test; $p < 0.01$), pointing to additional and critical information that appears to be derived from the combination of these two seemingly independent cues.

Fig. B.4 also shows results with inverted whole heads. As expected from the well-studied face-inversion effect (Yin, 1969), performance suffers in this condition relative to the upright head condition, and this gap between upright and inverted is observed for all viewing distances. Interestingly, however, the computed union of conditions B and C closely matches the empirically observed performance with inverted heads, suggesting the possibility that under inversion, the internal and external features act as independent contributors to identity processing. Their mutual configuration is utilized when the face is upright and leads to a significant boost in performance.

B.5 COMPUTATIONAL EXPERIMENTS

To probe whether the kinds of cue sensitivities that we observed in the data from human subjects could potentially emerge spontaneously through experience with full-face images, we have trained a computational model system – a deep convolutional neural network – on the task of face identification and subsequently tested its performance on various versions of face images, akin to the set-up presented in the human experiments.

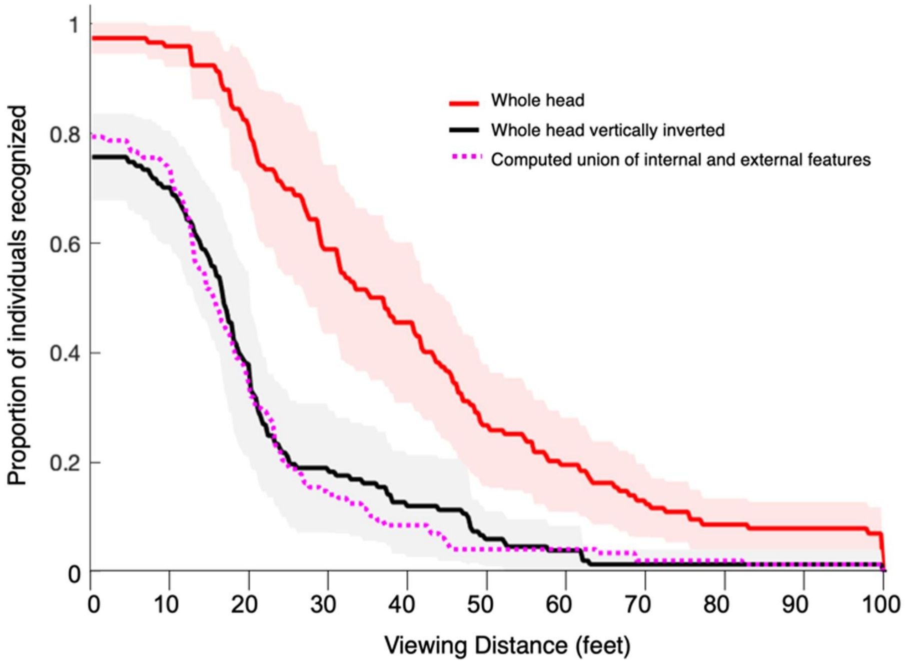


Figure B.4: Computed union of performances obtained in conditions B and C (dotted magenta line) relative to performance with whole heads (red line). Also shown is the performance with inverted whole heads (black line). The shaded regions around each line represent 95 % confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

B.5.1 *Methods*

The data used for training and testing our computational model were derived from the FaceScrub database (Ng and Winkler, 2014), a database containing images of mostly frontal celebrity faces, along with boundary box coordinates for generating tightly cropped versions thereof. As cropping with the default coordinates resulted in a partial cut-off of some of the outer facial features, we extended the width and height of the boundary boxes by 50 % on each side and disregarded all images where such extension was not possible. The resulting dataset, containing a total of 44,301 images, belonging to 524 different facial identities, was split into a 90 % training set and a 10 % test set. Of the 10 % test set, 100 clean and full-frontal exemplar images (belonging to 100 different faces) were selected and, akin to the stimulus preparation that was undertaken in the human experiments, manually edited in Photoshop, to produce three different test sets: the original full-face images, images in which only outer facial features were preserved, and images in which only inner facial features were preserved. The training images were fed into the AlexNet CNN (Krizhevsky et al., 2012), which was trained for 500 epochs using the stochastic gradient descent optimizer with a constant, low learning rate of 0.0001, Nesterov momentum of 0.9, and a batch size of 64. As part of preprocessing, normalization, random cropping (from 256 × 256 to 227 × 227 pixels), and random vertical flips were applied to the training data. Subsequently, the network's performance was plotted on the above-described test images consisting of three image conditions: full-face, outer-features only, and inner-features only. Finally, the union of the performances for inner-features only and outer-features only was computed (to compute the union, a given image was thereby evaluated as correctly classified if the inner-features only, the outer-features only images, or both were classified correctly).

To better understand the resulting performance patterns exhibited by our computational model system, and to try to enable a better comparability between the human and computational testing conditions, we conducted three additional control experiments. First, to reduce the impact of image-level modifications on the network's performance (considering that the removal of inner or outer facial features creates significant modifications to the overall image statistics), we added to the internal-features only versions of our 100 test images the external facial features of a computed 'average face' (the average face was obtained from the 'Face of Tomorrow' project conducted by South African photographer, Mike Mike, which involved averaging over 100 faces), thereby creating a more naturalistically-composed image without adding diagnostic information for a particular face image. Similarly, we added to the external-features only version of our 100 test images the internal facial features of the same 'average face' (see

Fig. B.5b and c for illustration). These new variations were then used to test the network again.

As part of a second control experiment, we sought to rule out the possibility that eventual performance decrements upon the selective elimination of either internal or external facial features would be the consequence of a mere lack of diagnostic information that the network draws upon prior to making a classification decision. This way, we hoped to better determine whether the network uses the internal and external facial information as independent cues to identity, or whether it uses additional information about their relative configuration, as appears to be the case in the human experiments. To achieve this, we tested the network's performance with images where the internal and external features were presented side-by-side, but not in natural spatial register, and compared it to the computed union of the performances with the two feature-sets shown as part of two separate images. In order to enable the network to process images with twice the width as the ones utilized in the previous computational experiment, we re-trained the network from scratch, on images where two copies of the exact same full-face were horizontally concatenated. Subsequently, we measured a) the classification performance when testing the network on horizontally-concatenated full-face images, b) the union of classification performances with horizontally-concatenated images containing only internal features and horizontally-concatenated images containing only external features, and c) the classification performance on horizontally-concatenated images containing both internal and external features that are positioned side-by-side (a side-by-side placement of Fig. B.5b and B.5c). In all cases, the internal or external features were presented along with the complementary (external or internal, respectively) 'average face' placeholders to avoid performance detriments due to significant modifications of the image statistics.

In our third, and final, control experiment, we examined the role of data augmentation on our network's performance patterns. Specifically, when considering the effect of blur to be a rough proxy for the effect of distance, it could, in principle, be argued that humans have experience not only with full-resolution facial images (i.e., when being close to another person's face) but with facial imagery which underwent an entire range of blur strengths. Thus, we here augmented the network's training set with whole face images across a range of resolutions, obtained by convolving the original images with Gaussians with standard-deviations of 0 through 5, and re-tested the performance on the different facial images used in the previous experiments.

B.5.2 *Results*

As expected, the trained network achieved high full-face classification performance, with an accuracy of 85 % on the overall test set contain-

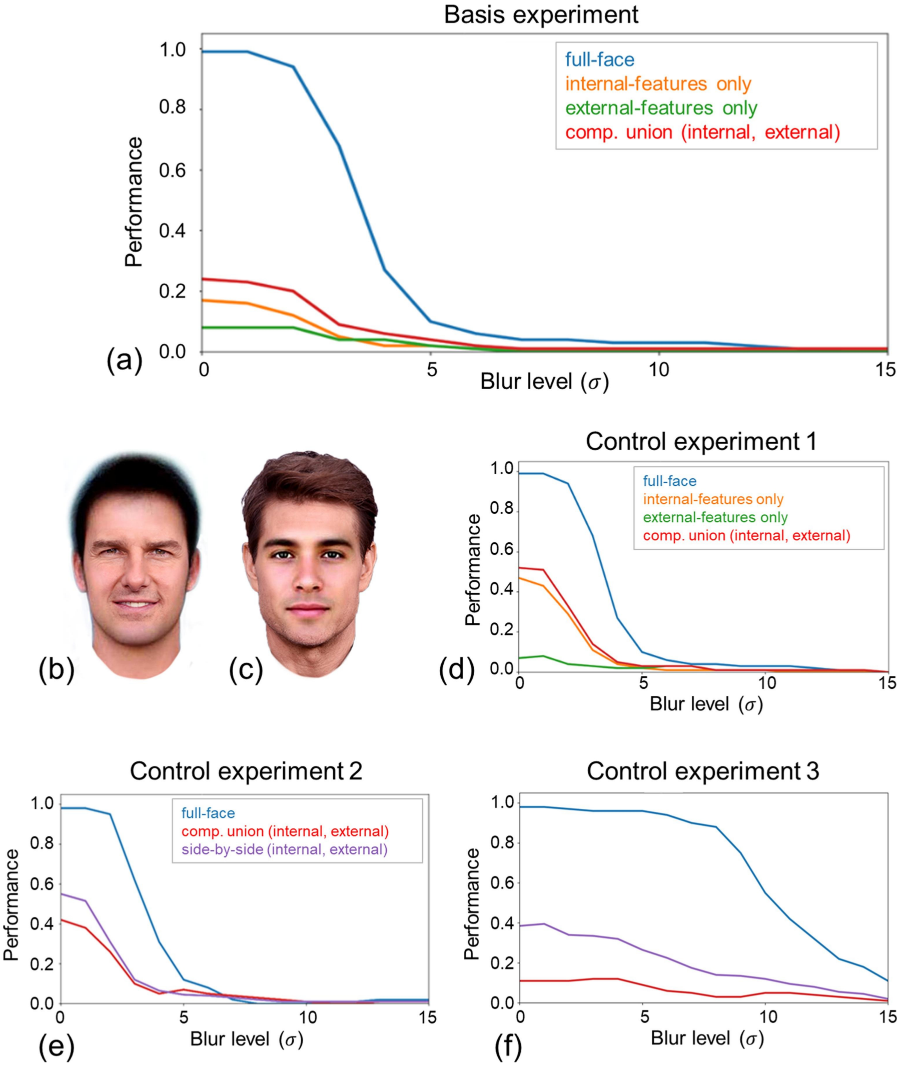


Figure B.5: Results from computational simulations. (a) Classification performances when training a network on full-face images and testing it on full-face, external-features only, and internal-features only images, as depicted in Fig. B.2. In addition, the computed union of internal-features only and external-features only performances is displayed (red line). (b) Exemplar of an image where only internal facial features are preserved and external features are filled in from an average face (average face obtained from the 'Face of Tomorrow' project conducted by South African photographer, Mike Mike, which involved averaging over 100 faces). (c) Exemplar of an image where only external facial features are preserved and internal features are filled in from an average face. (d) Classification performance when training a network on full-face images and testing it on the two image types shown in (b) and (c), along with full-faces. In addition, the computed union of internal-features only and external-features only performances is displayed (red line). (e) Classification performances when training a network on full-face images and testing it on full-face images (blue line), the computed union of internal-only and external-only performances (red line), and a side-by-side presentation of internal-only and external-only features (purple line). (f) Classification performances when training comprises not only high-resolution full-face images, but images convolved with Gaussian filters with standard-deviations ranging between 0 and 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ing 4430 images, and an even higher accuracy of 98 % on the subset of especially clean and full-frontal 100 test images used for all subsequent modifications and experiments. In contrast, when recognition performance was tested on the internal-features only or the external-features only test sets, the network's performance dropped markedly, even in the absence of any blur (see Fig. B.5a). The computed union of performances on the inner-only and outer-only test sets was, therefore, far below the classification performance achieved on full-face images. Thus, unlike in the human experiments, where the computed union was similar to full-face performance for low-blur settings but reached progressively lower performance levels when increasing the amount of blur, in the computational simulations reported here, given the network's low performance on the internal-only and external-only test sets, this pattern was evident across all resolution levels.

As introduced in the Methods section above, we carried out three control experiments to gain a better understanding of the nature of the performance differences between the human and computational system.

First, when adding external facial features of an 'average face' to the internal-features only test set and, similarly, adding internal facial features of the same 'average face' to the external-features only test set, the classification performances revealed an interesting differential effect: as shown in Fig. B.5d, while performance on the internal-features only test set increased markedly (from below 20 % to nearly 50 % for blur level 0), performance on the external-features only set remained unchanged (below 10 % for all blur levels). This points to a differential contribution of internal versus external facial features to overall classification performance of the deep neural network – one in which the network strongly favors internal features. Second, even when presenting internal and external features side-by-side, as part of the same image, classification performance levels remain markedly lower than the full-face presentation (see Fig. B.5e) and, instead, are comparable to their computed union, across all blur levels. Third, when augmenting the training set with whole face images across a range of blur levels (from 0 to 5), not surprisingly, performance as a function of resolution is improved (see Fig. B.5f). Interestingly, however, the inadequacy of the computed union of performance obtained with internal and external feature-sets separately, as also the much lower performance resulting from a side-by-side presentation of internal and external features, relative to that obtained with whole head images, are not less but even slightly more strongly evident in these data.

Taken together, these results suggest that although internal features yield higher classification performance than external ones, neither of the two feature-sets on their own, nor their computed union, are adequate to explain the performance of the network with full face images. This shortfall is observed whether the two kinds of cues are presented

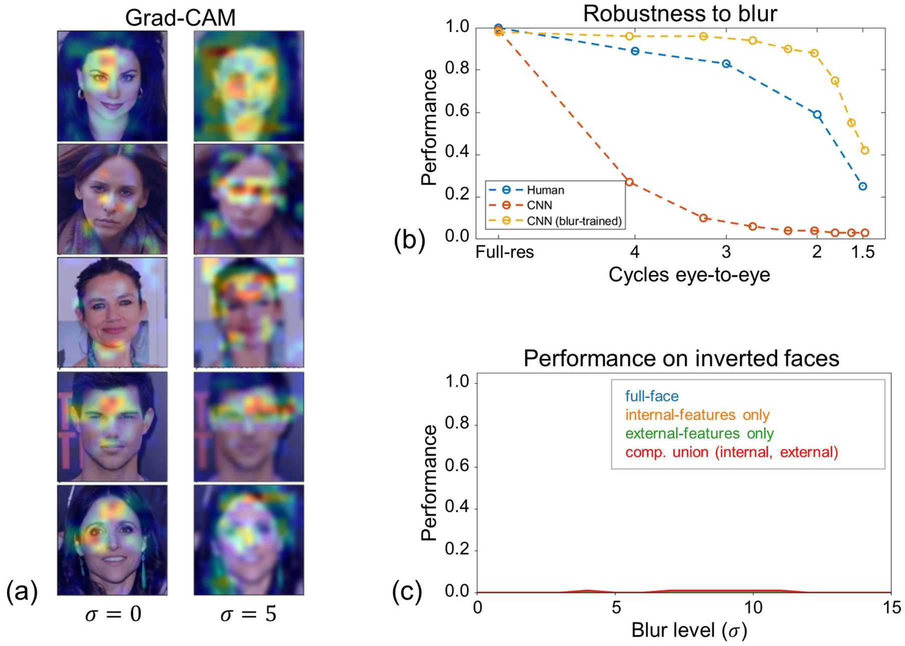


Figure B.6: (a) Gradient-weighted Class Activation Mapping (Grad-CAM) visualization applied to full-face images at blur levels 0 and 5, highlighting regions in the image that were important for the classification decision. (b) Human and CNN performance (trained only on full-resolution images, i.e., the 'basic experiment' setup reported earlier (see red line plot), as well as trained on mixed blur, i.e., the 'control experiment 3' setup reported earlier (see yellow line plot)) on full-faces as a function of blur. (c) CNN performance on upside-down faces, revealing chance performance across image type and blur level.

independently and their union computed later, or the two are shown simultaneously, but in non-natural spatial placement. In contrast to the human experiments, this shortfall does not gradually emerge with increasing blur levels but is evident even for high-resolution images.

While part of these results may be accounted for by the networks' generally greater fragility to image perturbations, beyond the control experiments we were able to carry out, it is possible that the network, rather than treating the internal and external features as independent cues to identity, may draw upon their mutual spatial configuration to arrive at an identity decision. Visualizing image regions important for the classification decision using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique (Selvaraju et al., 2017) reinforces this possibility. As shown in Fig. B.6a, for a blur level of 0, we see that only tiny regions of the face (mostly local, internal features) contribute to the decision. For a blur level of 5, we see an enlargement of the contributing regions, naturally coming to encompass external features.

B.6 DISCUSSION

The experiments we have described here allow us to make interesting inferences about the cues used for facial identification at different viewing distances, as well as the potential origin of these processes. First, our experiments characterize full-face identification performance as a function of available image resolution or distance. This kind of result has been previously reported in the literature by a few researchers, including Bachmann (1991), Costen et al. (1996), Harmon (1973), and Harmon and Julesz (1973). However, our results with whole heads address some important limitations of earlier studies. For instance, Harmon and Julesz's (1973) results of recognition performance with block-averaged images of familiar faces were confounded by the fact that subjects were told which of a small set of people they were going to be shown in the experiment. Later studies have also suffered from this problem. Bachmann (1991) and Costen et al. (1996) used a few high-resolution photographs during the 'training' session and low-resolution versions of the same photographs during the 'test' sessions. The subject's priming about stimulus set and the use of the same base photographs across the training and test sessions renders these experiments' results as limited for making inferences about real-world recognition. Another drawback of these studies that widens the gulf between the experiments and real-world settings, is that the images used were exclusively monochrome. Past experiments have shown that color increasingly contributes to object recognition with decreasing image resolution (Yip and Sinha, 2002). Therefore, we believe that the results reported here with full-color images are more representative of performance in real-world viewing conditions.

In addition to characterizing full-face recognition performance as a function of viewing distance, the more notable finding from this work is that exploring the distance dimension reveals a qualitatively different salience distribution across facial features from what has come to constitute the conventional view. More than the mutual configuration of the eyes, nose and mouth (Burton et al., 2005; Ellis et al., 1979; Hosie et al., 1988; Maurer et al., 2002; Valentine, 1991; Young et al., 1985; Yovel and Kanwisher, 2004), we find that it is the relationships between these features and external head contours that become especially crucial for recognition at increasing viewing distances. A summation of performances obtained with internal and external features separately cannot account for the performance obtained with the full face, in contrast to reports of independent contributions of the two feature sets (e.g., Betts and Wilson, 2010). This finding suggests that it is not the internal or external configurations on their own that subserve recognition, but rather measurements corresponding to how internal features are placed relative to the external features. This result complements brain-imaging reports showing modulation of face-selective neural responses by external features (Andrews et al., 2010; Axelrod and Yovel, 2010; Cox et al., 2004) and forces a reconsideration of conventional notions of facial configuration, which primarily involve 'internal' measurements such as inter-eye, eye to nose-tip and nose-tip to mouth distances (Brunelli and Poggio, 1993; Chen et al., 2001; Doi et al., 1998). What our results demonstrate is that the configural template deployed for recognition of faces at all but possibly the closest distances, necessarily incorporates measurements that link internal features with external ones. Additional support for this idea comes from the observations reported in Gilad-Gutnick et al. (2018). They find that independently transforming internal and external features disrupts recognition performance more significantly than transforming both together. The so-called 'Presidential illusion' (Gilad-Gutnick and Sinha, 2017; Sinha and Poggio, 1996, 2002) also serves to illustrate this idea, since it vividly demonstrates the insufficiency of internal features alone as determiners of identity.

Our finding of a greater reliance on external features for face recognition at low resolutions has an interesting analogue in the developmental literature. Reports from several researchers studying face recognition by neonates (Bushneil et al., 1989; Field et al., 1984; Pascalis et al., 1995) suggest that infants initially depend more on external features than on internal ones for discriminating between individuals. For instance, Pascalis et al. (1995) found that although four-day old infants could reliably discriminate their mother's face when all the facial information was present, they were unable to make the distinction when their mother and a stranger wore scarves around their heads. In conjunction with the fact that infant visual acuity starts out being very poor and improves over time (Dobson and Teller, 1978; Hainline and Abramov, 1992), these results echo our finding with adult observers. It is plausible, therefore, that infant reliance on external features may, as for our adult subjects, be driven at least in part by considerations of which subset of facial information provides more useful cues to identity at a given resolution.

The computational results presented in this paper furthermore suggest that the specific usage of facial features observed in humans could, in principle, emerge spontaneously through experience with full-face images. However, the congruence observed between human and CNN performance must be considered a partial and incomplete one for reasons alluded to above. While both systems appear to benefit from the simultaneous presence of internal and external feature sets in their correct spatial configuration, suggesting the potential usage of cross-set relationships, their absolute levels of performance vary strongly. CNNs are much worse with each feature set alone than humans. This may be the result of neural networks' generally greater fragility to image perturbations, beyond the dimensions we were able to control for, which, in turn, may, in part, account for the superiority of classification performance on full faces, relative to the computed union of their parts, even for the lowest blur level. The networks are also more vulnerable to performance decrement as a function of blur (see red line in Fig. B.6b) except if they are explicitly trained on blur, such as in control experiment 3 (see yellow line in Fig. B.6b), consistent with, for instance, Jang and Tong (2021) and Vogelsang et al. (2018). It is worth noting that in our simulations, the CNN is trained on all identities across a range of blurs. Humans appear to be able to transfer their resilience to blur across identities, i.e., even without having seen all identities in all different blur conditions. Further experimentation is needed to bring the artificial and biological training regimens in closer register.

We were unable to examine any relative changes in the contribution of facial attributes to overall classification when turning faces upside down as such examination, when not having trained the network on rotated faces as part of the preprocessing, resulted in chance-level performance – another notable difference to the performance of human participants (see Fig. B.6c). In addition to the partial congruence, these points of divergence could provide a future target for computational model systems in the ongoing effort of determining what aspects of biological vision and human experience need to be incorporated to replicate salient aspects of human performance. In addition to the use of more diverse, challenging, and ecologically-relevant databases and benchmarks, this could, among others, include probing the specific effect of training regimens incorporating aspects of human development (e.g., Vogelsang et al., 2018), specific training styles that may induce more global processing (e.g., Geirhos et al., 2018), or modifications of the task from supervised to unsupervised neural network training (e.g., Zhuang et al., 2021).

Further, considering the time-to-decision measurement procedure we utilized in experiment 2, we would like to note that there could possibly be an association between familiarity and time to recognition. Since our studies were conducted in a self-timed manner, with no extraneous time limits imposed, our data cannot speak to this association, but this issue would be a fruitful avenue for future inquiry. Further note that the starting distance in this experiment was 25 feet (since the facial images we used in the experiment had been scaled down from actual size, we were able to conduct the study within the physical constraints of a lab space).

Finally, we point out that we have not attempted to merge the results of experiments 1 and 2. Doing so requires establishing a mapping between viewing distance and effective image resolution available to an observer at that distance. This mapping cannot be determined simply from attributes like Snellen acuity, contrast sensitivity function, and photoreceptor density in the retina. It needs an empirical investigation. While we do not yet have this mapping, we are working on determining it and expect to report the results in a forthcoming paper.

In conclusion, by exploring the distance dimension and simulating the limiting conditions of face recognition, this study sheds new light on the relative significance of internal and external features to the demands of everyday face identification tasks. The pragmatic significance of such understanding lies in helping to design artificial recognition systems that may be better suited to dealing with the kinds of image degradations common to real settings.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Izzat N. Jarudi: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. Ainsley Braun: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. Marin Vogelsang: Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Visualization. Lukas Vogelsang: Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Visualization. Sharon Gilad-Gutnick: Writing – review & editing. Xavier Boix Bosch: Writing – review & editing. Walter V. Dixon: Writing – review & editing, Funding acquisition. Pawan Sinha: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

The authors wish to thank Neil Hazra and Sidharta Vadaparty. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2022-21102100009]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

Andrews, Timothy J, Jodie Davies-Thompson, Alan Kingstone, and Andrew W Young (2010). "Internal and external features of the face are represented holistically in face-selective regions of visual cortex." In: *Journal of Neuroscience* 30.9, pp. 3544–3552.

Axelrod, Vadim and Galit Yovel (2010). "External facial features modify the representation of internal facial features in the fusiform face area." In: *Neuroimage* 52.2, pp. 720–725.

Bachmann, Talis (1991). "Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity?" In: *European Journal of Cognitive Psychology* 3.1, pp. 87–103.

Betts, Lisa R and Hugh R Wilson (2010). "Heterogeneous structure in face-selective human occipito-temporal cortex." In: *Journal of Cognitive Neuroscience* 22.10, pp. 2276–2288.

Brunelli, Roberto and Tomaso Poggio (1993). "Face recognition: Features versus templates." In: *IEEE transactions on pattern analysis and machine intelligence* 15.10, pp. 1042–1052.

Burton, A Mike, Rob Jenkins, Peter JB Hancock, and David White (2005). "Robust representations for face recognition: The power of averages." In: *Cognitive psychology* 51.3, pp. 256–284.

Burton, A Mike, Stephen Wilson, Michelle Cowan, and Vicki Bruce (1999). "Face recognition in poor-quality video: Evidence from security surveillance." In: *Psychological Science* 10.3, pp. 243–248.

Bushneil, IWR, F Sai, and Jim T Mullin (1989). "Neonatal recognition of the mother's face." In: *British journal of developmental psychology* 7.1, pp. 3–15.

Chen, Li-Fen, Hong-Yuan Mark Liao, Ja-Chen Lin, and Chin-Chuan Han (2001). "Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof." In: *Pattern recognition* 34.7, pp. 1393–1403.

Cheung, Olivia S, Jennifer J Richler, Thomas J Palmeri, and Isabel Gauthier (2008). "Revisiting the role of spatial frequencies in the holistic processing of faces." In: *Journal of Experimental Psychology: Human Perception and Performance* 34.6, p. 1327.

Costen, Nicholas P, Denis M Parker, and Ian Craw (1996). "Effects of high-pass and low-pass spatial filtering on face identification." In: *Perception & psychophysics* 58, pp. 602–612.

Cox, David, Ethan Meyers, and Pawan Sinha (2004). "Contextually evoked object-specific responses in human visual cortex." In: *Science* 304.5667, pp. 115–117.

De Jong, Marloes, Willem A Wagenaar, Gezinus Wolters, and Ilse M Verstijnen (2005). "Familiar face recognition as a function of distance and illumination: A practical tool for use in the courtroom." In: *Psychology, Crime & Law* 11.1, pp. 87–97.

Dobson, Velma and Davida Y Teller (1978). "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies." In: *Vision research* 18.11, pp. 1469–1483.

Doi, M., K. Sato, and K. Chihara (1998). "A robust face identification against lighting fluctuation for lock control." In: *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 42–47.

Ellis, Hadyn D, John W Shepherd, and Graham M Davies (1979). "Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition." In: *Perception* 8.4, pp. 431–439.

Field, Tiffany M, Debra Cohen, Robert Garcia, and Reena Greenberg (1984). "Mother-stranger face discrimination by the newborn." In: *Infant Behavior and development* 7.1, pp. 19–25.

Fiorentini, Adriana, Lamberto Maffei, and Giulio Sandini (1983). "The role of high spatial frequencies in face perception." In: *Perception* 12.2, pp. 195–201.

Fotios, S, J Uttley, and S Fox (2018). "Exploring the nature of visual fixations on other pedestrians." In: *Lighting Research & Technology* 50.4, pp. 511–521.

Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel (2018). "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness." In: *arXiv preprint arXiv:1811.12231*.

Gilad-Gutnick, Sharon, Elia Samuel Harmatz, Kleovoulos Tsourides, Galit Yovel, and Pawan Sinha (2018). "Recognizing facial slivers." In: *Journal of cognitive neuroscience* 30.7, pp. 951–962.

Gilad-Gutnick, Sharon and Pawan Sinha (2017). "The presidential illusion." In: *The Oxford compendium of visual illusions*. Oxford University Press, pp. 628–632.

Goffaux, Valerie, Barbara Hault, Caroline Michel, Quoc C Vuong, and Bruno Rossion (2005). "The respective role of low and high spatial frequencies in supporting configural and featural processing of faces." In: *Perception* 34.1, pp. 77–86.

Greene, Ernest and Scott C Fraser (2002). "Observation distance and recognition of photographs of celebrities' faces." In: *Perceptual and motor skills* 95.2, pp. 637–651.

Hahn, Carina A, Alice J O'Toole, and P Jonathon Phillips (2016). "Dissecting the time course of person recognition in natural viewing environments." In: *British Journal of Psychology* 107.1, pp. 117–134.

Hainline, Louise and Israel Abramov (1992). "Assessing visual development: Is infant vision good enough." In: *Advances in infancy research*.

Hall, Edward T (1966). *The Hidden Dimension*. Anchor.

Harmon, Leon D (1973). "The recognition of faces." In: *Scientific American* 229.5, pp. 70–83.

Harmon, Leon D and Bela Julesz (1973). "Masking in visual recognition: Effects of two-dimensional filtered noise." In: *Science* 180.4091, pp. 1194–1197.

Hosie, Judith A, Hadyn D Ellis, and Nigel D Haig (1988). "The effect of feature displacement on the perception of well-known faces." In: *Perception* 17.4, pp. 461–474.

Jang, Hojin and Frank Tong (2021). "Convolutional neural networks trained with a developmental sequence of blurry to clear images reveal core differences between face and object processing." In: *Journal of vision* 21.12, pp. 6–6.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks." In: *Advances in neural information processing systems* 25.

Lampinen, James Michael, William Blake Erickson, Kara N Moore, and Aaron Hittson (2014). "Effects of distance on face recognition: Implications for eyewitness identification." In: *Psychonomic bulletin & review* 21, pp. 1489–1494.

Lewin, Catharina and Agneta Herlitz (2002). "Sex differences in face recognition—Women's faces make the difference." In: *Brain and cognition* 50.1, pp. 121–128.

Lindsay, RCL, Carolyn Semmler, Nathan Weber, Neil Brewer, and Marilyn R Lindsay (2008). "How variations in distance affect eyewitness reports and identification accuracy." In: *Law and Human Behavior* 32, pp. 526–535.

Loftus, Geoffrey R and Erin M Harley (2005). "Why is it easier to identify someone close than far away?" In: *Psychonomic Bulletin & Review* 12.1, pp. 43–65.

Maurer, Daphne, Richard Le Grand, and Catherine J Mondloch (2002). "The many faces of configural processing." In: *Trends in cognitive sciences* 6.6, pp. 255–260.

Megreya, Ahmed M and Markus Bindemann (2009). "Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect." In: *Perception* 38.12, pp. 1831–1848.

Mousavi, Seyed Morteza and Ipek Oruc (2020). "Size effects in the recognition of blurry faces." In: *Perception* 49.2, pp. 222–231.

Ng, Hong-Wei and Stefan Winkler (2014). "A data-driven approach to cleaning large face datasets." In: *2014 IEEE international conference on image processing (ICIP)*. IEEE, pp. 343–347.

O'Donnell, Christopher and Vicki Bruce (2001). "Familiarisation with faces selectively enhances sensitivity to changes made to the eyes." In: *Perception* 30.6, pp. 755–764.

Oliva, Aude, Antonio Torralba, and Philippe G Schyns (2006). "Hybrid images." In: *ACM Transactions on Graphics (TOG)* 25.3, pp. 527–532.

Oruc, Ipek, Fakhri Shafai, Shyam Murthy, Paula Lages, and Thais Ton (2019). "The adult face-diet: A naturalistic observation study." In: *Vision research* 157, pp. 222–229.

Parker, Denis M, J Roly Lishman, and Jim Hughes (1996). "Role of coarse and fine spatial information in face and object processing." In: *Journal of Experimental Psychology: Human Perception and Performance* 22.6, p. 1448.

Pascalis, Olivier, Scania de Schonen, John Morton, Christine Deruelle, and Marie Fabre-Grenet (1995). "Mother's face recognition by neonates: A replication and an extension." In: *Infant behavior and development* 18.1, pp. 79–85.

Ruiz-Soler, Marcos and Francesc S Beltran (2006). "Face perception: An integrative review of the role of spatial frequencies." In: *Psychological research* 70, pp. 273–292.

Schwaninger, Adrian, Stefan Ryf, and Franziska Hofer (2003). "Configural information is processed differently in perception and recognition of faces." In: *Vision Research* 43.14, pp. 1501–1505.

Selvaraju, Ramprasaath R, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization." In: *Proceedings of the IEEE international conference on computer vision*, pp. 618–626.

Sinha, Pawan and Tomaso Poggio (1996). "I think I know that face..." In: *Nature* 384.6608, pp. 404–404.

Sinha, Pawan and Tomaso Poggio (2002). "Last but Not Least." In: *Perception* 31.1. PMID: 11922119, pp. 133–133.

Valentine, Tim (1991). "A unified account of the effects of distinctiveness, inversion, and race in face recognition." In: *The Quarterly Journal of Experimental Psychology* 43.2, pp. 161–204.

Vogelsang, Lukas, Sharon Gilad-Gutnick, Evan Ehrenberg, Albert Yonas, Sidney Diamond, Richard Held, and Pawan Sinha (2018). "Potential downside of high initial visual acuity." In: *Proceedings of the National Academy of Sciences* 115.44, pp. 11333–11338.

Wagenaar, Willem A and Juliette H Van Der Schrier (1996). "Face recognition as a function of distance and illumination: A practical tool for use in the courtroom." In: *Psychology, Crime and Law* 2.4, pp. 321–332.

Wright, Daniel B and Benjamin Sladden (2003). "An own gender bias and the importance of hair in face recognition." In: *Acta psychologica* 114.1, pp. 101–114.

Yin, Robert K (1969). "Looking at upside-down faces." In: *Journal of experimental psychology* 81.1, p. 141.

Yip, Andrew W and Pawan Sinha (2002). "Contribution of color to face recognition." In: *Perception* 31.8, pp. 995–1003.

Young, Andrew W, Dennis C Hay, Kathryn H McWeeny, Brenda M Flude, and Andrew W Ellis (1985). "Matching familiar and unfamiliar faces on internal and external features." In: *Perception* 14.6, pp. 737–746.

Yovel, Galit and Nancy Kanwisher (2004). "Face perception: domain specific, not process specific." In: *Neuron* 44.5, pp. 889–898.

Zhuang, Chengxu, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C Frank, James J DiCarlo, and Daniel LK Yamins (2021). "Unsupervised neural network models of the ventral visual stream." In: *Proceedings of the National Academy of Sciences* 118.3, e2014196118.

C

ON PRENATAL AUDITORY EXPERIENCE IN HUMANS AND ITS RELEVANCE FOR MACHINE HEARING

Content from

Vogelsang, M.*, Vogelsang, L.*, Diamond, S., & Sinha, P. (2021). "On prenatal auditory experience in humans and its relevance for machine hearing". **Poster presented** at ICLR Workshop "General- ization beyond the training distribution in brains and machines", 2021, Online. (* = equal contribution)

C.1 ABSTRACT

Given the markedly better generalization capabilities of the human perceptual system relative to computational models, a question naturally arises about the genesis of this disparity. Here, we propose that a key to robust human perception might lie in its developmental trajectory. Unlike standard computational training procedures, perceptual development in humans undergoes a stereotypical temporal progression in which sensory inputs are initially highly degraded and gain quality later on. We focus here on the auditory domain, in which this progression commences already before birth: A fetus' experience in the womb comprises low-pass filtered versions of voices and other sounds in the environment. Such degraded inputs may induce the acquisition of mechanisms capable of performing extended temporal integration, facilitating robust analysis of information carried by slow variations in the auditory stream, such as emotions or other prosodic content. To computationally test this proposal, we assessed the consequences of training with different temporal progressions of filtered audio signals on a deep convolutional neural network's internal representations and subsequent classification of emotional prosodic content. We found that training with an auditory trajectory approximately mimicking the pre-to-post-natal progression yielded best generalization performance; it significantly exceeded outcomes following exclusively full-frequency, exclusively low-frequency, or inverse-developmental training protocols. The developmentally-trained model further acquired temporally extended receptive fields in its first convolutional layer and, when tested with fullfrequency inputs, exhibited the strongest resilience to the ablation of units tuned to high frequencies. These simulations suggest that the progression from low-tofull-frequency signals, rather than being an epiphenomenon, may be an enabling feature of perceptual development, conferring significant benefits to later auditory

processing. The results also point to the utility of incorporating similar procedures into the training of computational model systems and, more generally, to the inspiration that human development may provide towards the goal of achieving more robust generalization.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

`http://postcards.miede.de/`

Thank you very much for your feedback and contribution.

Final Version as of November 1, 2023 (classicthesis v4.6).

Erklärung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen.

Kapitel 2 – Vogelsang, Marin*; Vogelsang, Lukas*; Diamond, Sidney; & Sinha, Pawan. (2023). "Prenatal auditory experience and its sequelae". Veröffentlicht in *Developmental Science*, 26(1), e13278. (* = Diese Autoren haben gleichermaßen beigetragen).

- Konzeptualisierung: **Marin Vogelsang**, Lukas Vogelsang, Sidney Diamond, Pawan Sinha. *Alle Autoren trugen zur Formulierung der übergreifenden Forschungsziele der Studie bei. Marin Vogelsang und Lukas Vogelsang trugen darüber hinaus zur weiteren Entwicklung der Studie bei, indem sie die geeignete komputationale Methodik konzipierten.*
- Datenerhebung und Analyse: **Marin Vogelsang**, Lukas Vogelsang. *Marin Vogelsang führte alle komputationalen Simulationen und alle Analysen durch, über die in der Veröffentlichung berichtet werden. Lukas Vogelsang führte vorherige Pilotsimulationen durch, welche allerdings nicht im veröffentlichten Manuskript enthalten sind, und gab Feedback zur Methodik und Visualisierung.*
- Schreiben – Erster Entwurf: **Marin Vogelsang**, Lukas Vogelsang.
- Schreiben – Kommentare und Überarbeitung: **Marin Vogelsang**, Lukas Vogelsang, Sidney Diamond, Pawan Sinha.

Kapitel 3 – Vogelsang, Marin*; Vogelsang, Lukas*; Gupta, Priti*; Gandhi, Tapan; Shah, Pragya; Swami, Piyush; Gilad-Gutnick, Sharon; Ben-Ami, Shlomit; Diamond, Sidney; Ganesh, Suma; & Sinha, Pawan. (Eingereicht). "Impact of early visual experience on later usage of color cues". Im Peer-Review-Verfahren bei *Science*. (* = Diese Autoren haben gleichermaßen beigetragen)

- Konzeptualisierung: **Marin Vogelsang**, Lukas Vogelsang, Sidney Diamond, Pawan Sinha. *Diese Autoren formulierten die übergreifenden Forschungsziele der Arbeit im Kontext der "Adaptive Initial Degradation"-Hypothese.*
- Datenerhebung und Analyse: **Marin Vogelsang**, Lukas Vogelsang, Priti Gupta, Tapan Gandhi, Pragya Shah, Piyush Swami, Sharon Gilad-Gutnick, Shlomit Ben-Ami, Suma Ganesh, Pawan Sinha. *Marin Vogelsang führte alle Computersimulationen und alle komputationalen Analysen durch, über die in diesem Artikel berichtet werden. Lukas Vogelsang führte zuvor Pilotsimulationen durch, welche allerdings nicht im finalen Manuskript enthalten sind. Marin Vogelsang implementierte das Online-Experiment. Priti Gupta, Tapan Gandhi, Pragya Shah und Piyush Swami sammelten experimentelle Daten in Indien. Suma Ganesh führte die Behandlung und Charakterisierung der Patienten in Indien durch. Marin Vogelsang, Lukas Vogelsang, Priti Gupta, Sharon Gilad-Gutnick, Shlomit Ben-Ami, und Pawan Sinha trugen zur Analyse der experimentellen Daten aus Indien bei.*
- Schreiben – Erster Entwurf: **Marin Vogelsang**, Lukas Vogelsang, Priti Gupta, Sidney Diamond, Pawan Sinha.

- Schreiben – Kommentare und Überarbeitung: **Marin Vogelsang**, Lukas Vogelsang, Priti Gupta, Tapan Gandhi, Pragya Shah, Piyush Swami, Sharon Gilad-Gutnick, Shlomit Ben-Ami, Sidney Diamond, Suma Ganesh, Pawan Sinha.

Kapitel 4 – Vogelsang, Marin; Vogelsang, Lukas; Pipa, Gordon; Diamond, Sidney; & Sinha, Pawan. (Eingereicht). "On the origin of the parvo- and magnocellular division: potential role of developmental experience". Eingereichtes Manuskript.

- Konzeptualisierung: **Marin Vogelsang**, Lukas Vogelsang, Pawan Sinha. *Diese Autoren trugen zur Formulierung der übergreifenden Forschungsziele der Studie bei. Marin Vogelsang trug darüber hinaus zur weiteren Entwicklung der Studie bei, indem sie die geeignete komputationale Methodik konzipierten.*
- Datenerhebung und Analyse: **Marin Vogelsang**. *Marin Vogelsang führte alle Simulationen und alle komputationalen Analysen selbstständig durch.*
- Schreiben – Erster Entwurf: **Marin Vogelsang**, Lukas Vogelsang.
- Schreiben – Kommentare und Überarbeitung: **Marin Vogelsang**, Lukas Vogelsang, Gordon Pipa, Sidney Diamond, Pawan Sinha.

Kapitel 5 – Vogelsang, Lukas*; **Vogelsang, Marin***; Pipa, Gordon; Diamond, Sidney; & Sinha, Pawan. (Eingereicht). "Butterfly effects in perceptual development: a review of the 'adaptive initial degradation' hypothesis". Im Peer-Review-Verfahren bei *Developmental Review*. (* = Diese Autoren haben gleichermaßen beigetragen)

- Konzeptualisierung: Lukas Vogelsang, **Marin Vogelsang**, Gordon Pipa, Sidney Diamond, Pawan Sinha.
- Datenerhebung und Analyse: Da es sich um ein review-paper handelt, wurden keine neuen empirischen oder komputationalen Daten erhoben oder analysiert.
- Schreiben – Erster Entwurf: Lukas Vogelsang, **Marin Vogelsang**.
- Schreiben – Kommentare und Überarbeitung: Lukas Vogelsang, **Marin Vogelsang**, Gordon Pipa, Sidney Diamond, Pawan Sinha.

Kapitel 6 – Gupta, Priti; Shah, Pragya; Gilad-Gutnick, Sharon; **Vogelsang, Marin**; Vogelsang, Lukas; Tiwari, Kashish; Gandhi, Tapan; Ganesh, Suma; & Sinha, Pawan. (2022). "Development of visual memory capacity following early-onset and extended blindness". Veröffentlicht in *Psychological Science*, 33(6), 847-858.

- Konzeptualisierung: Priti Gupta, Sharon Gilad-Gutnick, Pawan Sinha. *Priti Gupta und Pawan Sinha entwickelten das Studienkonzept. Priti Gupta, Sharon Gilad-Gutnick und Pawan Sinha konzipierten die Experimente.*
- Datenerhebung und Analyse: Priti Gupta, Pragya Shah; **Marin Vogelsang**, Lukas Vogelsang, Kashish Tiwari, Guma Ganesh, Pawan Sinha. *Priti Gupta, Pragya Shah und Kashish Tiwari führten die Experimente durch, und Priti Gupta, Pragya Shah, Lukas Vogelsang, und Pawan Sinha analysierten die Daten. Marin Vogelsang führte die Computersimulationen- und analysen durch, zu denen Lukas Vogelsang durch Feedback beitrug. Suma Ganesh führte die operativen Verfahren und die Charakterisierung der Patienten durch.*
- Schreiben – Erster Entwurf: Priti Gupta, Pragya Shah, Sharon Gilad-Gutnick, **Marin Vogelsang**, Lukas Vogelsang, Tapan Gandhi, Pawan Sinha.
- Schreiben – Kommentare und Überarbeitung: Priti Gupta, Pragya Shah, Sharon Gilad-Gutnick, **Marin Vogelsang**, Lukas Vogelsang, Kashish Tiwari, Tapan Gandhi, Suma Ganesh, Pawan Sinha.

Kapitel 7 – Bi, Shakila; Chawariya, Ajay; Ganesh, Suma; Gupta, Priti; Huang, Youqi; Jazayeri, Kimiya; Kumar, Rakesh; Ralekar, Chetan; Singh, Chaitanya; Tiwary, Adya; Vogelsang, Lukas; **Vogelsang, Marin**; Yadav, Mrinalini; & Sinha, Pawan. (2023). "Scholastic status of congenitally blind children following sight surgery". Veröffentlicht in *International Journal of Special Education*, 37(2), 160-168. (Alphabetische Autorenreihenfolge mit Ausnahme von Pawan Sinha)

- Konzeptualisierung: Pawan Sinha.
- Datenerhebung und Analyse: Shakila Bi, Ajay Chawariya, Suma Ganesh, Priti Gupta, Youqi Huang, Kimiya Jazayeri, Rakesh Kumar, Chetan Ralekar, Chaitanya Singh, Adya Tiwary, Lukas Vogelsang, **Marin Vogelsang**, Mrinalini Yadav, Pawan Sinha.
- Schreiben – Erster Entwurf: Shakila Bi, Ajay Chawariya, Suma Ganesh, Priti Gupta, Youqi Huang, Kimiya Jazayeri, Rakesh Kumar, Chetan Ralekar, Chaitanya Singh, Adya Tiwary, Lukas Vogelsang, **Marin Vogelsang**, Mrinalini Yadav, Pawan Sinha.
- Schreiben – Kommentare und Überarbeitung: Shakila Bi, Ajay Chawariya, Suma Ganesh, Priti Gupta, Youqi Huang, Kimiya Jazayeri, Rakesh Kumar, Chetan Ralekar, Chaitanya Singh, Adya Tiwary, Lukas Vogelsang, **Marin Vogelsang**, Mrinalini Yadav, Pawan Sinha.

Appendix A – Gupta, Priti; Shah, Pragya; Gilad-Gutnick, Sharon; **Vogelsang, Marin**; Vogelsang, Lukas; & Sinha, Pawan. (Eingereicht). "The influence of semantics on visual memory capacity in children and adults". In Revision bei *British Journal of Developmental Psychology*.

- Konzeptualisierung: Priti Gupta, Sharon Gilad-Gutnick, Pawan Sinha. *Priti Gupta und Pawan Sinha entwickelten das Studienkonzept. Priti Gupta, Sharon Gilad-Gutnick und Pawan Sinha konzipierten die Experimente.*
- Datenerhebung und Analyse: Priti Gupta, Pragya Shah; **Marin Vogelsang**, Lukas Vogelsang, Pawan Sinha. *Priti Gupta und Pragya Shah führten die Experimente durch, und Priti Gupta, Pragya Shah, Lukas Vogelsang, und Pawan Sinha analysierten die Daten. Marin Vogelsang führte alle Computersimulationen- und analysen durch.*
- Schreiben – Erster Entwurf: Priti Gupta, Pragya Shah, Sharon Gilad-Gutnick, **Marin Vogelsang**, Lukas Vogelsang, Pawan Sinha.
- Schreiben – Kommentare und Überarbeitung: Priti Gupta, Pragya Shah, Sharon Gilad-Gutnick, **Marin Vogelsang**, Lukas Vogelsang, Pawan Sinha.

Appendix B – Jarudi, Izzat; Braun, Ainsley; **Vogelsang, Marin**; Vogelsang, Lukas; Gilad-Gutnick Sharon; Bosch, Xavier Boix; Dixon, Walter; & Sinha, Pawan. (2023). "Recognizing distant faces". Veröffentlicht in *Vision Research*, 205, 108184.

- Konzeptualisierung: Izzat Jarudi, Ainsley Braun, Pawan Sinha.
- Datenerhebung und Analyse: Izzat Jarudi, Ainsley Braun, **Marin Vogelsang**, Lukas Vogelsang, Pawan Sinha. *Izzat Jarudi und Ainsley Braun sammelten die experimentellen Daten und analysierten diese zusammen mit Pawan Sinha. Marin Vogelsang bereitete die Stimuli für die komputationale Studienkomponente vor und führte alle Computersimulationen durch. Marin Vogelsang und Lukas Vogelsang führten komputationale Analysen durch.*
- Schreiben – Erster Entwurf: Izzat Jarudi, Ainsley Braun, **Marin Vogelsang**, Lukas Vogelsang, Pawan Sinha.
- Schreiben – Kommentare und Überarbeitung: Izzat Jarudi, Ainsley Braun, **Marin Vogelsang**, Lukas Vogelsang, Sharon Gilad-Gutnick, Xavier Boix Bosch, Walter Dixon, Pawan Sinha.

Andere Kapitel: Alle anderen Kapitel wurden ausschließlich von mir, Marin Vogelsang, verfasst. Lediglich geringfügiges sprachliches Feedback wurde im Rahmen des Probelesens von Kapitelteilen eingearbeitet von Priti Gupta (Englisch), Jannis Born (Englisch), und Lukas Vogelsang (English/Deutsch). Darüber hinaus wurde DeepL zur Unterstützung und Kontrolle der Übersetzung vom Englischen ins Deutsche verwendet.

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

....
(Ort. Datum) (Unterschrift)