A Predictive Processing Theory of Autism: A Neural Network Modelling Approach with Applications to Autistic Savantism

Beren Millidge

Master of Science
Artificial Intelligence
School of Informatics
University of Edinburgh
2017

Abstract

This dissertation presents a novel framework for understanding autistic spectrum disorders (ASD) within the predictive processing paradigm. It argues that the observed pattern of long-range underconnectivity and local overconnectivity observed in ASD, when instantiated in a predictive processing framework, will lead to impoverished high level regions and lower-level regions which down-weight prior information coming from above and prioritise incoming immediate sensory information. This pattern is hypothesised to result in many behaviours recognised in autism such as sensory hypersensitivity, a local, detailed-oriented processing style, and a preference towards predictable behaviours and routines. Neural network models were constructed to investigate several aspects of this theory. We showed that the autistic networks were superior at discriminative tasks and inferior at integrative tasks, which matches findings in the autism literature. We also investigated the effect of this connectivity pattern on hemisphericity and found that networks with a hemispheric split with lesser cross-connectivity exhibited more autistic behaviours and some also showed signs of developing autistic savantism. These results, then, offer a novel theory of autism from within the predictive processing paradigm and modelling evidence in favour of the theory which may expand our understanding of the disorder.

Acknowledgements

I would like to thank my supervisor, Dr. Richard Shillcock for his guidance throughout the project, as well as my co-supervisees Nick Johnson, Yuansheng Song, and Katy Fukou for many enlightening and informative conversations. I would also like to thank Florian Bolenz, an MSc student with Richard last year, for making available code which forms the basis of the work on autistic savantism in this dissertation.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Beren Millidge*)

Table of Contents

1 Introduction **1**

2 Background **3**

2.1 Autism . . . 3

2.2 Theories of Autism . . . 5

2.2.1 Theory of Mind . . . 5

2.2.2 Weak Central Coherence . . . 5

2.2.3 Neurobiological Explanations . . . 7

2.2.4 Synaptic Pruning . . . 9

2.2.5 Predictive Processing . . . 10

2.3 Integrated Framework . . . 12

2.4 Autistic Savantism . . . 15

2.5 Related Work . . . 17

3 Experiments **19**

3.1 Artificial Neural Networks . . . 19

3.2 Predictive Processing Models . . . 21

3.3 Discriminator Networks . . . 21

3.3.1 Methodology . . . 21

3.3.2 Results . . . 23

3.4 Integrative networks . . . 29

3.4.1 Methodology . . . 29

3.4.2 Results . . . 31

3.5 Autistic Savantism . . . 35

3.5.1 Methodology . . . 35

3.5.2 Results . . . 37

4 Discussion **45**

4.0.1 Limitations 49

4.0.2 Future Work 51

4.1 Conclusion 52

Bibliography **53**

Chapter 1

Introduction

Autism is a neurodevelopmental disorder characterised by a triad of impairments in social interaction, impairments in communication, and repetitive and obsesssive interests. It is also associated with a number of perceptual and motor abnormalities. A number of theories of autism have been proposed, including the Theory of Mind Theory [1], the Weak Central Coherence Theory [2], and the Cortical Underconnectivity theory. In this dissertation we propose and develop a theory of autism within the predictive processing paradigm and construct several different models to provide an empirical grounding for our theory.

Predictive processing is a general cognitive paradigm proposed by Friston et al [3]. In brief, it states that the brain functions by continuously attempting to predict incoming sensorimotor contingencies, and then adapts its internal structure to improve its ability to make accurate predictions. The brain, according to Friston, is composed of a hierarchical series of multi-level generative probabilistic models which are capable of modelling various aspects of the external world and also using the models to sample predictions of how the world will be. It is these predictions that are compared to the incoming sensory information at every step, and the models are adjusted to minimise the prediction error. Information flows up and down the hierarchy. The generative model at each level receives sensory information flowing up and predictions flowing down the hierarchy, integrates them into a united posterior belief, and then minimises the difference between this posterior and the next set of sensorimotor information flowing up from the hierarchy. In this way the brain as a whole becomes able to predict, and thus understand, its environment.

In this paper we propose a novel predictive processing theory of autism which is based on neurophysiological data. Many recent fMRI and other brain imaging techniques

have revealed that neural connectivity in autism differs systematically from the neurotypical type. In short, individuals with autism appear to have systemic long range under-connectivity, and may have short range over-connectivity. We propose that this connectivity pattern, when instantiated in a predictive processing model will naturally result in many of the deficits observed in autism. This pattern of long-range under-connectivity and short range over-connectivity may also generalise to a hemispheric pattern of lesser interhemispheric connectivity in autism. We hypothesise that this hemispheric pattern may lead to a greater variance in the cognitive skills of those with autism, and may also aid the development of savant skills in some individuals.

We construct neural network models to test these hypotheses. We find that predictive processing models with impaired long range connectivity perform slightly better at a sensory discrimination task, and are significantly more sensitive to small peturbations of stimuli, which matches the psychophysical findings in autism. We also show that such networks are poorer at integrating stimuli from different modalities together, a key finding in autism and a cornerstone of many theories of autism such as the Temporal Binding Deficit theory [4]. We also show that networks with a hemispheric split obtain more variance in training when they possess less inter-hemispheric connectivity and also sometimes develop savant skills, although the exact mechanism behind this remains somewhat mysterious.

The dissertation is structured as follows: In chapter 2, we present the historical and intellectual background of the theoretical and neurophysiological work on autism. We explore theories of autism as well as what recent decades of fMRI and other data tell us about the neural underpinnings of the disorder. We also present our own predictive processing approach. In the third chapter, we construct models which implement our predictive processing model of autism and conduct experiments to confirm, albeit in a simplified and abstract model, the predictions made by the theory. We also construct a hemispheric model of autism and conduct experiments on the variance and emergence of savant like skills in such a model. In the fourth chapter, we analyse these results, and discuss their limitations, and future work that could be done in this area.

Chapter 2

Background

2.1 Autism

Autism is a pervasive developmental neurological condition that manifests as an array of deficits in areas such as social cognition and communication [1], rigid and obsessive interests, insistence on sameness and routine [5], and a detail-oriented processing style [6]. Autism is often defined through the triad of impaired social interaction, impaired communication, and repetitive and obsessive interests [7]. However, many individuals with autism also possess a wide range of perceptual and motor abnormalities including both hypo and hyper sensitivity, and a relative lack of susceptibility to many common sensory illusions [8], [9].

First identified by Kanner (1943) [10] and Asperger (1944) [11], it is thought to affect about 6 in 1000 children [12] and incidence has been increasing in recent times in developed countries. [13]. This increase is likely due to changing diagnostic criteria and better mental health screening in the developed world rather than an increase in the true frequency [14]. The symptoms and behaviours characteristic of autism have remained remarkably stable since the initial child case-studies of Kanner and Asperger in the early 20th century. Autism ranges in its degree of severity, from relatively high functioning conditions such as Asperger's syndrome, to low functioning autism and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). Because of this the DSM V reclassified the cluster of similar conditions as simply being autistic-spectrum disorders (ASD).

Autism has a substantial genetic component. The heritability of autism have been es-

timated to be as high as 90% due to monozygotic twin concordance rates between 73-95% [15]. Other studies have found a lower, but still substantial heritability of about 50% [16] with a significant effect of shared environment. The sibling risk stands at about 20% for males and 10% for females. The risk for a broader, non-diagnosable autistic phenotype, manifesting as aloof or tactless speech, obsessive interests, need for routine, and a difficulty forming friendships may be significantly higher [17]. Given such high heritability, it is unsurprising that many possible genetic loci for autism have been found. More than 100 candidate genes have been analysed for possible linkages with autistic spectrum disorders [18]. Some of the most promising are the serotonin transporter gene 5-HTT, although evidence here is mixed, with contrary studies such as Ramoz et al (2006) claiming no such linkage [19]. The neuroligin gene [20] has also been considered. Other possible pathways involve the mTOR/PI3K pathway, mutations to which might cause abnormal synaptic growth rate in autism [21], [22] and may interact with various mutations in pathways involving serotonergic system [23]. Mutations in the NLGN3/4 or NRXN1 have also been hypothesised to alter synaptic function in a way that might induce autism or mental retardation [24].

Tracing the mechanism by which alterations at these genetic loci can lead to the macro-level deficits and abnormal neural development observed in autism is a difficult task, as there is vast number of possible genetic pathways as more than one third of the human genome is expressed at some point in the developing brain [25], and interactions between these pathways as well as environmental influences may have a significant impact. Moreover, many initial studies positing linkages fail to replicate [13].

In this paper we shall primarily focus on higher level theories of autism and ASD, which attempt to explain the disorder at a computational, algorithmic, or functional level. The precise mechanisms by which genetic abnormalities generate such macro-scale differences remain to be elucidated.

2.2 Theories of Autism

2.2.1 Theory of Mind

Early theories of autism focused primarily on explaining the social deficits. One paradigm is the Theory of Mind theory, proposed by Baron-Cohen (1985) [1]. He posits that autistic children lack a consistent Theory of Mind - a metarepresentational capacity [26] to understand that other agents have beliefs and desires and intentions just as they do, and to be able to understand their behaviours in terms of those mental states and to be able to use that knowledge to make inferences about the mental states of others [27]. Evidence for this lack includes the failure of autistic children on the Sally-Anne test [1], [28] even when compared to IQ-matched controls and developmentally delayed children with Downs syndrome. Further evidence comes from the lack of pretend play in autistic children [29].

However, as more evidence came to light of the large and seemingly disparate number of non-social impairments and deficits in ASD, then such social theories slowly became untenable. While it might be possible to describe difficulties in communication and interaction as being caused by a lack of theory of mind, increased perceptual sensitivity and lesser habituation to simple visual stimuli [30] are much more difficult to explain in such a framework. It is thus becoming increasingly clear that autism is a systemic disorder rather than one constrained to a single module of the brain. If autistic patients lack a theory of mind, then it is an effect of a broader systemic deficit.

Many new theories have emerged seeking to explain the patterns of dysfunction and altered-function in ASD as the result of some systemic deficit. Rather than focusing on a single dysfunctional subsystem or module such Theory of Mind or Executive Dysfunction [31] these theories instead postulate an abstract dysfunction on a general computational or algorithmic level which, when instantiated in the brain, can explain the pattern of seemingly unrelated deficits as are observed in ASD.

2.2.2 Weak Central Coherence

Perhaps the most important and widespread of these theories is the Weak Central Coherence Theory of Frith and Happe [2] [32]. This theory argues that instead of being

caused by a specific impairment, ASD is characterised by a certain processing style called Weak Central Coherence. Neurotypical individuals generally try to bind together the information they receive and search for gestalts and abstract concepts to explain away the detail. Autistic individuals, on the other hand, tend not to utilise more abstract integrative processing styles, but instead tend to focus more purely on the intricacies of the input without feeling such a need to tie it all together. Whereas the processing style in normal subjects can be thought of as centralised and hierarchical, the style in ASD is much more local and distributed; it lacks central coherence to a much greater extent than normal.

Weak Central Coherence Theory can explain why individuals with autism often outperform normal controls in many perceptual tasks which require detailed attention to local features such as the Embedded Figures Task [33]. Moreover, autistic individuals often tend to preferentially process parts of stimuli as shown by Plaisted et al (1999) [34], who presented them with hierarchical stimuli - such as a big letter composed of many smaller letters. They found that the ASD group performed better at rapid judgements of the identity of the low-level letters than the controls, while the controls did better at integrating the global percept to determine the identity of the high-level letter.

Similarly, individuals with ASD have been found to be helped less by global organisation of information as shown in Jarrold and Russell (1997) [35] who presented subjects with patterns of dots to count. The dots were shown either in a random pattern or a 'canonical' pattern which matched that found on the faces of a die. Autistic subjects showed little improvement from the canonical presentation whereas controls showed a significant improvement, implying that individuals with ASD either are unable to utilise, or do not require, the aid given by the globally coherent organisation of information. Subjects with ASD also often perform poorly on tasks which require the global integration of information such as integrating fragments of images [36].

Although the weak central coherence theory can be used to explain a wide range of disparate deficits in ASD, the core construct of the theory 'central coherence' is vague and underdefined. It is not clear whether the deficit lies in a single coordinating coherence module, or arises organically as the result of a broader, systemic abnormality. In their original paper, Frith and Happe (1994) appear to suggest the former. On the other hand, the temporal binding deficit theory states that individuals with ASD are impaired

at binding together and integrating percepts from different parts of the brain [4], and that this causes the general pattern of behaviour characterised by weak central coherence. This theory suffers from the same problem as weak central coherence theory in that both theories focus primarily on the computational and representational levels of Marr's hierarchy of explanation [37]. They postulate abstract deficits in central coherence or temporal binding, but leave the underlying mechanism which causes such a deficit mostly unspecified. Ideally a complete theory of autism would possess both a computational component (what deficits exist and what effects they have in abstract terms) and an implementational component (what exactly is physically wrong or different about the neural hardware so as to cause such deficits). There are a number of theories which propose explanations for ASD situated at the neurobiological level.

2.2.3 Neurobiological Explanations

One of the most promising of these is the theory of cortical underconnectivity which has emerged from recent fMRI data. In the past decade, many findings of cortical and functional underconnectivity between different brain regions have been observed in autism. For instance, underconnectivity has been reported between brain regions in social and emotional tasks [38], [39], global processing and cognitive control [40], [41], working memory [42], theory of mind [43], and visuospatial attention [44].

Functional underconnectivity has also been reported in the resting state [45], [46], where subjects are instructed to lie in the fMRI scanner, but not to think of anything in particular. In these tasks underconnectivity was typically found between the task-specific posterior areas and the pre-frontal cortex, which is where the inputs of these different inputs are thought to be integrated and synthesised.

Underconnectivity has also been found in direct region to region connections outside of the fronto-posterior network, such as between the primary and supplementary motor areas and the cerebellum and thalamus [47], between the fusiform gyurs and the amygdala [48], between the visual cortex and the thalamus [49], and between the anterior cingulate and the frontal eye fields [50].

In addition to information about functional connectivity obtained through fMRI, converging evidence comes from diffusion tensor imaging indicates that children with

ASD possess abnormal white matter distributions indicative of a disrupted pattern of structural connectivity [51], [52].

Moreover, there is circumstantial evidence that this functional underconnectivity might be causal. Several studies have investigated whether there is a relationship between the degree of underconnectivity and the severity of ASD. Just et al (2007) [42] found a positive correlation between functional underconnectivity and scores on the Autism Diagnostic Observation Schedule (ADOS), a set of semi-structured psychiatric tests used to asses severity of ASD. Similarly, Monk et al (2009) [53] has found that poorer social functioning (as measured by the Autism Diagnostic Interview-Revised) is correlated with weaker functional connectivity between the superior frontal gyrus and posterior cingulate cortex.

Although these studies are correlational, so that the observed correlation between degree of autism severity and functional under-connectivity can theoretically be explained by a third latent factor underpinning them both, it nevertheless gives some stronger evidence towards the hypothesis that functional under-connectivity and ASD are causally related than the previous studies which simply reported the co-occurrence of the two effects.

Functional overconnectivity has also been reported in ASD, for instance in the extrastriate (visual) cortex [54], the amygdala [55], and the parahippocampal gyri [56]. Nair et al (2013) [57] found overconnectivity in temporo-thalamic regions while Monk et al (2009) [53] reported higher connectivity in the posterior cingulate cortex. Bailey et al (1998) [58] also found overconnectivity in the cerebello-thalamo cortical pathway, which they attribute to reduced numbers of inhibitory Purkinje cells. There have also been a few studies done on the relationship between degree of overconnectivity and the severity of ASD symptoms. Gusnard et al (2009) [53] found that in adolescents with ASD, overconnectivity in the medial prefrontal cortex and parahippocampal gyrus correlated with poorer verbal and nonverbal skills. Moreoever, Agam et al (2010) [50], found that higher connectivity within the anterior singulate and frontal eye fields was correlated with restricted, stereotyped behavioural patterns reminiscent of autism. The same caveats apply to these correlational studies as apply to the underconnectivity ones discussed previously.

Although there are several contrary studies, including ones which report mixtures of over and underconnectivity within the same regions [59], [60], the general pattern appears to be one of widespread cortical underconnectivity with the underconnectivity especially pronounced between relatively distant brain regions - and local overconnectivity [61], and that the autistic brain may rely primarily on local connectivity rather than long-range connectivity for the transmission of information [62].

2.2.4 Synaptic Pruning

To understand what effects this pattern of aberrant connectivity might have on brain function, we need to consider the brain in terms of functional specialisation and functional integration [3]. To accomplish the multitude of functions that the brain must complete in normal operation, it requires specialised brain regions to deal with various tasks such as sensory processing of different modalities, planning, memory, and so forth (functional specialisation). However, for the brain to function as a whole and behave in a cohesive manner, different functional regions need inputs from other regions and information must be combined effectively and integrated throughout the brain. Thus the brain needs a significant degree of functional integration as well as specialisation [63]. The theory of cortical under-connectivity argues that in ASD this complex equilibrium is distorted so that overall functioning is impaired. Specifically, it hypothesises that the deficits in long range connectivity between regions damages the functional integration of the brain. This would explain the local-processing bias observed in ASD and explained by theories such as Weak Central Coherence as well as serve as a neurobiological underpinning for impaired temporal binding between regions, as proposed by the Temporal Binding Deficit theory.

Although the mechanism by which this abnormal pattern of connectivity arises in autism is not well understood, one plausible hypothesis is that it is the result of the abnormal developmental trajectory observed in the first few years of life of autistic infants. This trajectory begins with a significant neural overgrowth, and perhaps ends with excessive synaptic pruning. Early brain development in neurotypical infants comprises a period of rapid brain growth and increase in white matter, followed by a period of plateauing and synaptic pruning where, it is thought, the weaker and unused connections are pruned back [64] [65]. Both of these phases seem more extreme in autistic

infants. At birth they have larger heads, on average, than neurotypical infants, and a significantly greater incidence of macrocephaly (head size in the 99th percentile) - between 10-30% of autistic infants are macrocephalic at birth [66], [67]. For the first few years of life, there is also evidence of brain overgrowth in autism [68], [69], of which a significant component is white matter. This suggests that the autistic infants brain may be significantly overconnected compared to that of a neurotypical infant.

There is also evidence that the synaptic pruning following brain overgrowth may be more aggressive in individuals with ASD, and that the pruning might disproportionately affect the long-range cross-region synapses [70]. If this is the case, then the observed developmental trajectory of overgrowth followed by overaggressive pruning would explain the abnormal connectivity patterns of long range underconnectivity and short range overconnectivity found in ASD, as the aggressive pruning would have eliminated most of the long-range connections, but the short range connections are less effected, so a shadow of the initial overgrowth remains.

2.2.5 Predictive Processing

In recent times, several Bayesian or predictive processing treatments of neurodevelopmental disorders like autism have been proposed [71], [72]. Predictive processing posits that instead of passively receiving and extracting statistical information from the sensory signal, the brain is constantly engaged in a process of trying to predict and infer the immediate future states of the world. In this way it solves what is known as the inverse problem which is that the brain must infer what is out there in the world merely from the sensory stimulations that it creates. This is an underconstrained problem so there are many equally valid solutions. To solve it, therefore, the brain must also incorporate some kind of prior knowledge or expectations about the world into its predictions and representations about the world state.

From a computational perspective, the predictive processing hypothesis is that the brain is organised into a series of hierarchical layers of probabilistic generative models [73]. Each generative model receives a set of inputs from the layer below it, and predictions propagate downwards from the layer above. The model then integrates these two sources of information to form a posterior belief about the world, which it can then

translate into predictions of expected future states. It then propagates these predictions down to the layer below. In such an arrangement, increasingly abstract and processed sensory information is propagated up the hierarchy while increasingly more specific and detailed predictions are propagated down to the lower levels [74]. At each level a process akin to Bayesian reasoning takes place as each layer receives upwards input (the likelihood), top-down predictions (the prior), and combines them to form its own model of the world (the posterior). After every set of inputs, the generative model is adjusted so as to iteratively update the generated predictions with respect to the true reality of what occurred. This update is done through the minimisation of the prediction error of that level which is simply the divergence between the prediction and the observed reality. In this way the brain encodes a set of dynamic causal models at every level of abstraction which can be expected to reflect the structure of the world, and which it can utilise to make predictions or effective action [75].

Due to the minimization of prediction error at every level, the brain can be considered a system which acts and perceives in such a way to minimise its surprisal, the gap between what it expects and what it perceives. Mathematically it does this by minimising the free energy functional [76], which means that from the space of all functions from sensory data to predictions (i.e. models), it will attempt find the function that minimises the Kullback-Leibler divergence between the distributions over sensory inputs and the distribution over predictions [77].

The models at the lowest level are the least abstract and are concerned almost entirely with trying to predict the low level sensory inputs at every time step. However, higher level models become increasingly wide-ranging and abstract, and are able to take into account prior knowledge, contextual information, or information obtained through other modalities and propagate this knowledge back down the hierarchy as priors and predictions to help modulate processing there. In this way extra-receptive field modulation effects, - such as boundary effects in V1 cells - can be explained. [78].

Several models have been proposed to try to explain autism spectrum disorders within the predictive processing framework. Perhaps the first was Pellicano et al (2012) [72] who argued that many behavioural symptoms of ASD such as sensitivity to tiny variations in stimulus, poor generalisation ability, and difficulty with complex tasks requiring the integration of contextual or temporally distant information might be the result

of attenuated priors. These hypo-priors mean that each layer is much less affected by the predictions of higher layers than expected, which implies that it is more difficult for the generative model to incorporate the top down information it needs to make accurate predictions.

Sinha et al 2014 [79] argue that autism can be explained as a disorder of prediction in which the high level generative models are unable to suitably model the world, thus leading the autistic individual to experience a fundamentally chaotic and confusing world with constant sensory overload due. Similarly, Van der Cruys et al (2014 [80] propose that the deficit in autism is in learning to suppress or ignore errors. Since the autistic individual is unable to do this effectively, they are thus overwhelmed with the number of small prediction errors they have made about the world, leading to feelings of sensory overload.

Another related hypothesis is that of Lawson et al (2014) [71] who argue that the cause of autism is that ASD individuals have an aberrant precision in the predictive processing model in that they tend to weight their incoming sensory evidence much more strongly than optimal which leads them to discount prior information coming from above. In a predictive processing model, this has essentially the same effect as the attenuated priors proposed by Pellicano et al, since the precisions of the prior and the sensory evidence are in direct competition such that increasing one must decrease the other.

2.3 Integrated Framework

In this paper we propose a novel framework which integrates insights from the Weak Central Coherence theory and the well-established aberrant patterns of functional and structural connectivity in ASD with the predictive processing paradigm.

Consider the effect of the disturbed pattern of connectivity found in autism - long range underconnectivity and short-range overconnectivity - on a predictive processing model consisting of a hierarchical sequence of probabilistic generative models. Each step in the hierarchy will generally involve a long range connection between different brain regions except, perhaps, at the lower levels which involve more tightly integrated sen-

sory regions. Because of the long-range underconnectivity, the bandwidth for such long range communication is much smaller than expected. This impairs the ability of sensory information to propagate up the hierarchy, and prior predictions to propagate down.

Because communication is somewhat impaired between regions, this means that the predictions (priors) propagated from each layer to the one below it are less impactful, more noisy, and possibly incomplete. This has the effect of attenutating the strength of the prior on the processing occurring at that level thus providing a solid neurophysiological mechanism for the attenuated priors postulated in Pellicano et al 2012 [72]. Moreover, as the priors are noisy and probably incomplete, this means that they are less predictive and useful than they should be in helping the layer generate correct predictions in the future. Thus, if the system is adaptive, it will learn to downgrade the importance of the priors compared to the sensory information since they are less informative than in neurotypicals. This explains Lawsons 2014 [71] argument in favour of a down-weighted precision on the priors.

The next thing to consider is the effect of the underconnectivity on the information flowing up the system. Since, at many steps, the connectivity between the adjacent regions in the hierarchy will be poor, information will have difficulty flowing up the hierarchy as well. This may have the effect of impoverishing the high-level abstract generative models at the top of the system to such an extent that they might become unable to predict the rapidly changing dynamic world of social interactions or face to face communications, leading to difficulties there. Moreover, such impoverished models, especially when beset by troubled communications, may also be unable to successfully integrate information converging there from multiple regions successfully, leading to a weak executive function and central coherence. In addition, if the higher levels are relatively unsuccessful, this means that the prior predictions deriving from these models which do propagate down to lower levels will be a fairly poor predictor of future sensory information coming up, and thus each layer will have an additional incentive to down-weight the precision of the prior predictions compared to the sensory stimuli, thus further biasing the overall characteristics of the network towards local processing.

The fact that the higher-level regions may become relatively impoverished and unable to respond successfully to complex situations may explain why many autistic indi-

viduals experience uncomfortable sensory overload in complex situations. It may also explain why they find social interaction and communication challenging social interactions require dealing with a rapidly changing dynamic situation, as well as integrating many subtle and ambiguous cues in real time to try to infer the internal mental states of others - a challenging task even for neurotypicals. Moreover, the common preference of individuals with ASD for simple, rigid rigid routines, as well as interests in mechanical systems which obey understandable rule sets may be a reflection of the relatively impoverished nature of the high level models. Since they are unable to model complex environments, but can successfully predict the behaviour of simpler mechanical objects and routinised behaviours, interacting obsessively with these simpler, and easier to understand systems, and undertaking only stereotyped and routinised behaviour may provide individuals with ASD with some relief compared to the cacophany of prediction errors they experience in more complex situations. In addition, many individuals with ASD may be drawn to stimming behaviours [81] which are simple, repetitive, predictable movements for similar reasons.

The effect of this connectivity pattern on the predictive processing model can also be considered through the lens of over and underfitting. In short, we argue that such a connectivity pattern would cause the lower levels of the hierarchy, close to the sensory input, to overfit, while the higher levels would underfit the data. This is because the lower levels are in immediate contact with the sensory region so their input is not disrupted by the abnormal connectivity pattern. Moreoever, the effect of the top-down priors on these regions is attenuated. The imposition of priors on a system has been shown to generally have a regularising effect and thus reduces overfitting. For instance, it has been shown that the effect of assuming a Gaussian noise prior on the weights of an artificial neural network is mathematically equivalent to that of L2 regularisation, a common and effective regularisation method in the Machine Learning community [82]. Since there is also an overabundance of local connectivity, this will also have the effect of increasing the degree of overfitting, since the expressive power of the region will increase without any concomitant increase of regularisation.

The fact that under this model it seems likely that low level regions overfit is interesting since many authors have commented on the similarity of several symptoms of ASD such as the sensitivity to minute differences in the input, the relative lack of generalisation ability and confusion when presented with similar but subtly different stimuli,

and the lack of habituation to repeated exposures to overfitting in neural networks. Indeed, one of the first neural network models of autism Cohen et al (1994) [83] modelled autism as overfitting in an associator network. Additional evidence comes from Plaisted et al (2015) [84] who argues, completely independently of any work on predictive processing, that there is reduced generalisation ability in ASD. It is also hypothesised that due to the relatively impoverished and incomplete data the higher levels receive due to the poor long-range connectivity, the higher levels are substantially under-fit, which renders them insufficient to model rapidly changing and complex environmental dynamics and integrate many sources of information together at once to guide behaviour and perception.

2.4 Autistic Savantism

A related question relates the functional pattern of long range underconnectivity and overconnectivity to hemispheric differences in autistic individuals. There have been multiple findings of reduced corpus callosal (the corpus callosum is the white matter tract which connects the two hemispheres of the brain) size in autism [85], [86]. This is in line with the more general pattern of long range underconnectivity. It is hypothesised that this weakened connectivity between hemispheres might lead to a greater variance in intelligence and skill acquisition in those with autism, and might even aid in the development of savant skills in some individuals with ASD.

This variance is thought to arise as a natural effect of the pattern of connectivity between hemispheres. Since the connectivity, and thus communication, between the hemispheres is decreased, they should operate and learn more autonomously. This autonomy may enable them to develop strategies and tools for processing the input more independently than otherwise. When required to coordinate across hemispheres, on some occasions, the tools that each hemisphere have learned will complement each other, thus performing the task well. While on other occasions they will fail to complement each other, and thus performance will be poor. Together, these effects create a greater variance in performance than when the hemispheres can communicate more closely. Mathematically, this can be modelled by the simple fact that the variance of the composition of two independent - or less correlated variables - will be greater than that of two highly correlated variables.

A similar explanation has been proposed to explain differences in intelligence between the sexes, as found in the Lothian Birth Cohort [87]. They found that while there is little significant differences in the means of the two distribution, men are typically higher in variance, leading to male over-representation at the tails of the distribution [88]. Ingalhalikar et al 2014 [89], argue that this might be due to neurophysiological sex differences in functional connectivity. They find that male brains are more modular and more specialised for intra-hemispheric connectivity while female brains are, on average, more widely connected and possess a greater degree of inter-hemispheric connectivity.

Shillcock et al (2016) [90] argues that some of this greater variance might be explained by the neuroscientific fact that men typically have smaller corpus callosums than women, relative to brain volume [91], although this is disputed [92], and that this might cause the greater degree of variance of intelligence. Bolenz et al (2016) [93], a previous masters student at the University of Edinburgh, model this phenomenon by using a network split into two hemispheres with a variable degree of connectivity between them which they use to produce ”male” and ”female” networks which exhibit the same pattern of variance of performance as real male-female distributions on IQ test data.

Since a similar connectivity pattern has been observed in autism as well, we adapt the model of Bolenz et al to our investigation of autism. We hypothesise that the increased variance might also lead to the development of savant skills in some individuals. Savant skills are those in which the autistic individual performs at a high level relative to their poor level of general functioning. Sometimes these skills can even surpass those of neurotypicals and even experts in that skill. Only a small subset of autistic individuals, estimated at around 10% of the autistic population appear to develop these skills [94], [95]. The skills are typically of a mnemonic or mathematical bent. Mathematical skills include calendrical calculations, prime number finding, and arithmetical operations such as the multiplication and division of large numbers. Other skills rely on an exceptional mnemonic ability and these include memorising railway timetables, musical pieces, or aerial views of cities. Despite their prodigious skill in some narrow domain, savants often have poor functioning in other areas, and thus contradict the positive manifold found in studies of general intelligence in which all aspects of intelligence correlate positively with each other [96]. Savant skills may develop because,

since the two hemispheres are less connected, each may become more autonomously, and can specialise more easily in a single skill. If one does choose to specialise, then it will be able to achieve a high performance in that skill, while sacrificing performance in the others, thus resulting in savant-like behaviour.

2.5 Related Work

Relatively little work has been done on neural network modelling of autism. This is largely because most theoretical work has been at a high level, and most empirical work has concentrated on finding neurobiological or phyisiological evidence supporting or refuting high level theses. Nevertheless, some work has been done.

Cohen 1994 [83] provides a neural network model which argues that children with autism simply have too many neurons and too much connectivity, so that their brain effectively overfits on the data they are exposed to, leading to excellent discrimination in known domains, but poor generalisation. Cohen's model varies the number of connections in the neural network on a fairly simple learning task. Networks with too few connections - which underfit- fare poorly on both the training and validation tests. Networks with too many connections - which overfit - do very well in the training phase but poorly in the test phase. Cohen argues that this is a feature that might explain some of the specific impairments and behavioural and perceptual abnormalities in autism. This linkage of overfitting with symptoms of autism is very interesting, however the neural network modelling actually undertaken here is mostly trivial, since it merely recapitulates that neural networks which are overparametrised tend to overfit - a fact already well known in the literature.

A similar mechanism is modelled by Vidal et al (2006) [97] who argue that the excessive brain-growth typically seen in early childhood of those with autism causes problems with the general integration and coordination of different brain regions, thus supporting the weak central coherence theory. This approach is also supported by neurophysiological evidence such as Stoner et al (2014) [98] who find that there are patches of disorganisation in the neocortex of people with autism which could be symptomatic of a rapid and uncontrolled early brain growth.

Thomas et al (2011) [70] propose another model in which it is not the gross connec-

tivity that causes problems in autism, but the much more severe synaptic pruning that typically follows it. They have built a model in which performance regressions occur, as is sometimes seen in autism, and in the more severe childhood disintegrative disorder (CDD) as a result of excessive pruning of connections which occurs after the growth. They implement this through increasingly severe pruning - or setting to zero - of weights within the model. As the amount of connections pruned increases, performance decreases, on average in the population of weights trained in the model. Some networks, in the population, moreover show the same pattern of regression as is observed in Childhood Disintegrative Disorder. They choose which connections to prune by setting a threshold weight below which the connection is pruned to zero. To model more aggressive pruning, they increase the threshold weight. It is unknown to what extent this relatively simple model of pruning can be generalised to synaptic pruning in the brain.

There are also a number of predictive processing models of various phenomena. Spratling et al (2015) provides an overview of various predictive processing algorithms. In addition, Rao and Ballard et al (1999) [78] build a predictive processing model of the lower visual cortex which can generate such phenomena as endstopping and context modulation in extrastriate visual neurons, while Pezzulo et al (2013) [99] provide a predictive processing approach to model-based reinforcement leaning. However, as far as we know, no work has been done on autism.

Our predictive processing model of autism, is thus, as far as we can tell, the first implemented predictive processing model for this disorder. It is thus hoped that it brings some small contribution to the available literature on autism and predictive processing. Similarly, no models have focused on the effects of hemisphericity and interhemispheric connectivity on autism and autistic savantism and thus, if successful, this project could lead to a significant increase in knowledge about this possibility.

Chapter 3

Experiments

3.1 Artificial Neural Networks

The modelling paradigm used throughout this dissertation is that of artificial neural networks. These arose early within the discipline of cognitive science, and still largely implement the follow of the McCullough and Pitts Neuron [100]. They became widely used in the connectionist paradigm of the 1980s and 1990s, and today form the basis of many modern machine learning methods.

The basis of artificial neural networks is the artificial neuron. Unlike the biological neuron, it implements a simple mathematical function. It receives a set of inputs. It sums the inputs and then passes the result through a non-linear function. This result of this process is the output, or 'activation' of the neuron. Each of the inputs is composed of the activation of another neuron in a previous layer, multiplied by a scalar weight which signifies the connection strength between the two neurons.

A single neuron can be represented mathematically as:

$$y = f(\sum_i w_i x_i + b)$$

Where y is the final output (activation) of the neuron. f() is a nonlinear function of the input. ws signify weights and xs signify inputs from previous neurons. b is called the bias term, since it enables the equivalent of a constant input or 'bias' to the neuron.

These neurons are arranged into layers. Since in artificial neural networks each neuron in the layer has no lateral connections with others in its layer, then each neuron is independent of the others in its layer. This means that the neurons in the layer can be concatenated into a single matrix, allowing the computation of the whole layer to be

done in a single matrix operation:

$$Y = f(Wx + B)$$

The activation function f must be non-linear. This is because any composition of linear functions is itself a linear function, and thus adding additional layers to the network would not add expressive power since the composition of many linear layers can be expressed identically in a single linear layer. As the function is non-linear, adding additional layers allows the network to learn deep abstractions which are built up hierarchically from the data. Intriguingly, this behaviour mimics that of the brain, where increasingly abstract representations are typically found in the deeper levels.

A neural network learns by adjusting the weight matrices between layers so as to match the output layer for any input with the desired output layer. There have been a number of methods proposed to achieve this. One of the most powerful and most general is that of gradient descent using backpropagation of error.

At each trial, the input is presented to the network and activations are propagated forward through the network to the output layer. The resulting output is then compared to the desired output in a way specified by a cost function. The gradient of the cost function can then be calculated with respect to the weights of the output layer. To minimise the cost function, the weights are adjusted in the direction which will decrease the cost function most rapidly. This direction is always perpendicular to the vector of the gradients of the weights. This method of learning produces the gradient descent learning rule, which can be expressed mathematically as:

$$\Delta w = -\eta \nabla w$$

where Δw is the change in weight value. η is a scalar coefficient called the learning rate and ∇w is the vector of partial derivatives of the weights which compose the gradient. If every operation in the neural network architecture is differentiable, then the gradient of the cost function with respect to every pair of weights in the network can be calculated and updated in such a fashion, thus optimising the whole network to the cost function simultaneously.

For all networks in this paper, a simple gradient descent optimiser was used for training. Unless specified otherwise, all layers used the sigmoid activation function.

3.2 Predictive Processing Models

To test our hypotheses we implement several predictive processing models. Although the work of Friston et al assumes a probabilistic setting, we find that maintaining full probability distributions over the space of possible input and output distributions is not mathematically tractable for large scale problems, so we collapse all distributions to their maximum a-posterior (MAP) points, where the MAP is operationalised as the pattern of activations in the output layer of the neural network representing the MAP point of the N-dimensional posterior distribution. This approach trades off the theoretical advantages of the fully Bayesian approach for an easily-implemented mathematically tractable model.

Since instead of using the full prediction and input distributions we are instead using MAP point estimates, then to minimise the prediction error we simply minimise the difference between the two sets patterns, rather than the Kullback-Leibler divergence between the two distributions, as is done in Friston et al.

Our cost function to minimise, then, is simply:

$$L = |y - p|$$

Where L is the cost function to be minimised, y is the posterior of the layer, and p is the top down prediction. This cost function is minimised via gradient descent. A precision weighting can also be set which weights the values of the inputs and predictions against each other.

3.3 Discriminator Networks

3.3.1 Methodology

Many studies have shown that autistic individuals often demonstrate superior perceptual abilities than controls. These abilities are especially manifest in tasks such as the Embedded Figures Task where local processing and attention to small, precise, details are the main demands [101]. We hypothesise that this is due to the impaired ability of top down predictions to be propagated down the sensory hierarchy, leading to low level sensory regions acting more autonomously. While this provides an exceptional sensitivity to fine detail, it also leads to impaired generalisation ability and a difficulty cohering information into gestalts.

To test this hypothesis we trained a predictive processing model to discriminate between similar stimuli. For these stimuli we used MNIST digits. MNIST is a dataset of grayscale 28 by 28 pixel images of single, handwritten, numbers from 0 to 9. Each number is centred in the image. Although long superseded by more difficult tasks, it has long served as a benchmark for computer vision in the machine learning community.

Although usually the MNIST dataset is utilised for classification, we designed a slightly different task - discrimination. Given two MNIST digits, the network must learn to determine whether they represent the same digit or different digits. Given that single digits can be written in a number of ways, this is not a trivial task to solve. We utilised this discrimination paradigm since we wanted to test the superior sensitivity - in effect the ability to discriminate - to sensory stimuli in autistic subjects.

The input stimuli are the two digits the network must discriminate between. Each input is presented to separate networks and processed independently for two layers. The layers are then combined in a difference layer that subtracts the activation of each of the two layers. This difference layer then projects to a single output unit. The activations of this unit are forced to lie between 0 and 1 through the use of a sigmoid activation function, so that it can be interpreted as a probability. A simple diagram of the network architecture is presented below:

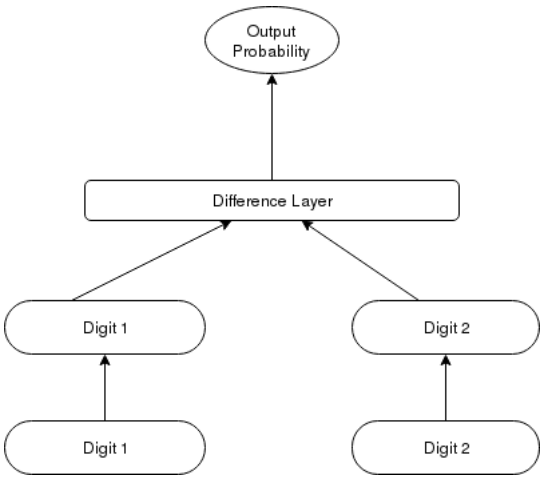

Figure 3.1: The network architecture of the discriminator network. Each of the two digits to be compared are fed through two layers of independent processing, their activations are then combined into a difference layer which is then used to compute the probability of the two digits being the same.

To simulate the long range underconnectivity and short range overconnectivity observed in autism, a weight mask was applied across all weight matrices in the autistic case, which randomly zeroed out some of the weights. This reduced the effective connectivity between the different layers of the network. The proportion of connections zeroed out was set at 10%. The neurotypical network had no weight matrix applied.

Our main hypothesis is that the autistic networks will be able to successfully learn to discriminate finer differences in the input stimuli than the control networks.

3.3.2 Results

The ability of the discriminative networks to discriminate between different digits was tested first. The training curves are shown below:

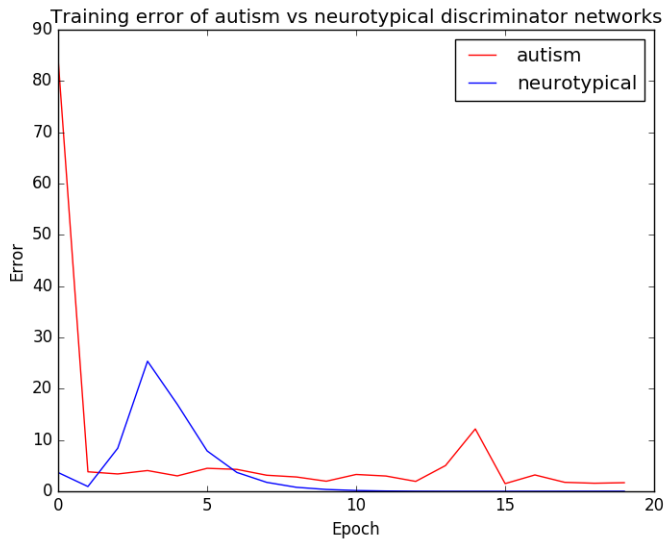

Figure 3.2: The error of the discriminative networks during training for both the autistic and neurotypical network

Both the autistic and the control networks perform significantly above chance by the end of training. Interestingly the autistic network performs slightly worse overall. Although it rapidly decreases its cost initially, it then plateaus at a significant cost while the non-autistic network improves beyond it. This goes against the initial hypothesis that autistic networks ought to have better performance, in line with the sensory advantages studies have found autistic subjects to possess on certain perceptual tasks. However, it is not a serious flaw since no study has claimed that subjects with ASD have superior perception overall than controls, which is what is modelled here.

The training accuracy over time of the two discriminator networks was also compared. A graph can be seen below:

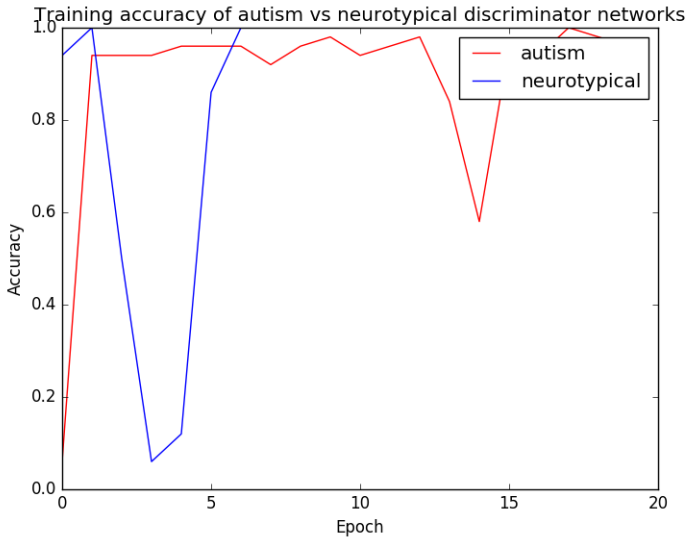

Figure 3.3: The accuracy of the discriminative networks during training for both the autistic and neurotypical network

Here a similar pattern emerges. Although the autistic network performs better initially, it later plateaus while the non autistic network goes on to achieve better performance overall.

The networks were tested again on validation data to ensure that the observed performance during training would also generalise to unseen data. The plot of the validation accuracy of the autistic and neurotypical network is shown below:

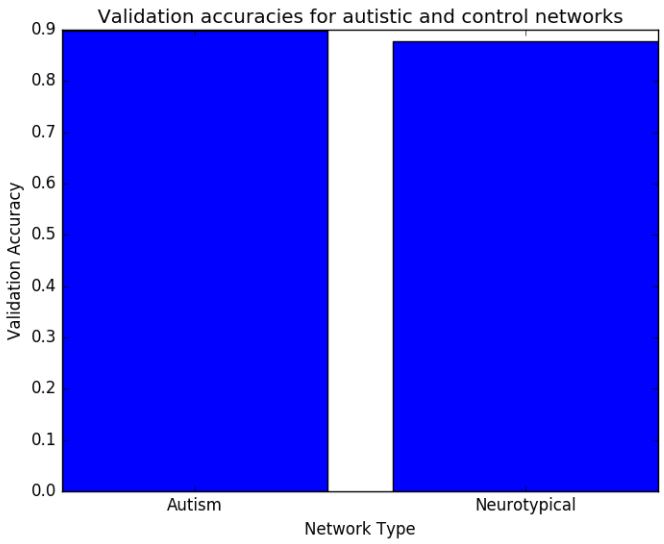

Figure 3.4: The accuracy of the discriminative networks during training for both the autistic and neurotypical network

Unlike in training, the autistic network performs slightly better on the validation data. Whether this is a significant effect, however, is uncertain, since the difference is minor. Both networks perform significantly above chance, demonstrating that this task can be successfully learnt by networks with both autistic and neurotypical connectivity patterns.

We then tested the ability of the network to discriminate fine differences between stimuli of the same digit.

The output of the network is a single scalar value - between 0 and 1 - which is interpreted as the probability that the two inputs are different. We set a threshold of 0.5, meaning that if the output was greater than 0.5, it was assumed that the network meant that the images were of different numbers, and less than 0.5 indicating that the images were of the same numbers. To create slightly different images we applied either a translation or a rotation to the original image. Examples of a translated and two rotated images compared to a standard are shown below:

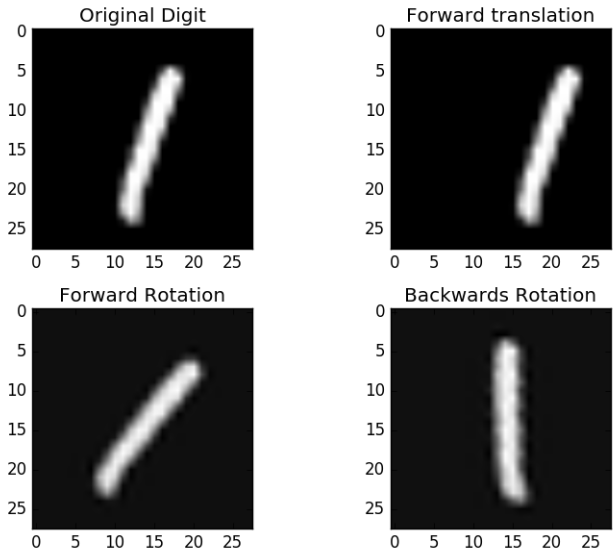

Figure 3.5: The accuracy of the discriminative networks during training for both the autistic and neurotypical network

It was hypothesised that the autistic networks would exhibit poorer generalisation and appreciation of gestalts than the control networks. This is instantiated as having greater discriminative ability so that an autistic network is significantly more likely to rate the slightly altered digit as different to the baseline digit than the control network. This was tested for different degrees of translation such as a pixel shift of between one and six pixels, and with rotation angles between -20 and +20 degrees.

The results below for the effect of translation of the original image are plotted below:

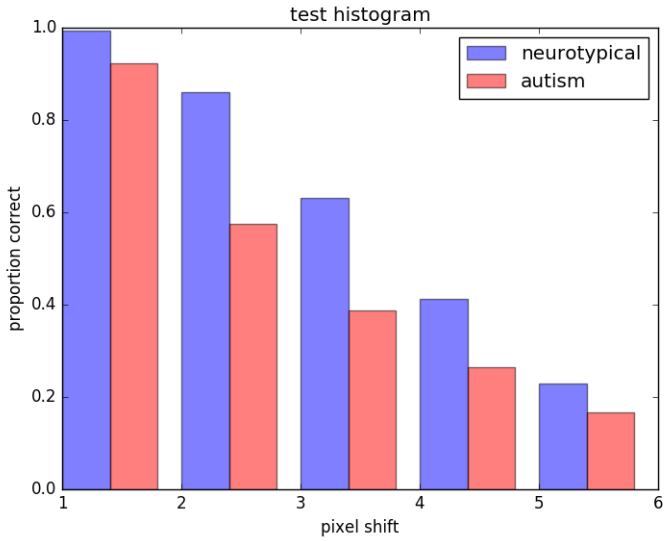

Figure 3.6: The proportion of correct responses identifying a translated digit with a non-translated digit, by the degree of translation, for the autistic and neurotypical discriminative networks

And for the rotation:

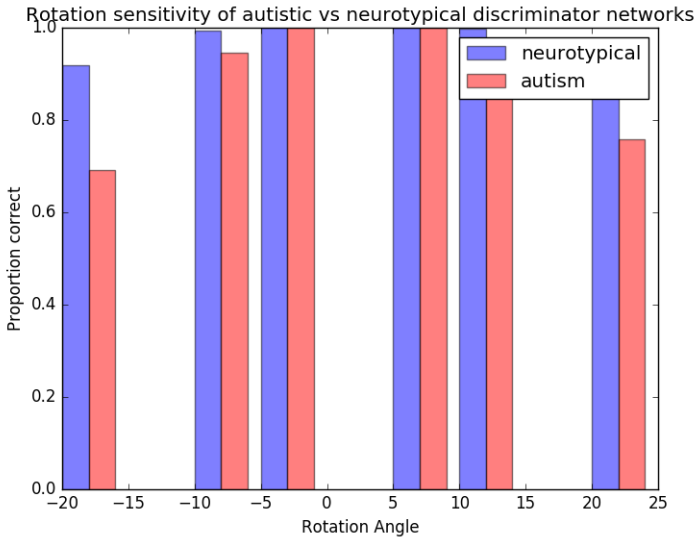

Figure 3.7: The proportion of correct responses identifying a rotated digit with a non-translated digit, by the angle of rotation, for the autistic and neurotypical discriminative networks

Since the identity of the digits are the same, just transformed, the percent correct on the y axis indicates how often the networks (corectly) claimed they were the same.

Although the accuracy of both networks declined with increasing degree of shift, the autistic networks were significantly poorer at this task which implies that they are more sensitive to small peturbations in the input stimuli, as predicted by the hypothesis.

This means that the autistic networks are significantly more sensitive to small shifts in the input stimuli, supporting a more local processing style as opposed to an integrated processing style allowing for abstraction, generalisation, and gestalts. This is broadly in line with empirical results showing that subjects with ASD exhibit superior discrimination results on tasks requiring primarily local information, as well as poorer generalisation. Moreover, it demonstrates that the hypothesised aberrant pattern of long range underconnectivity can lead to networks which exhibit such ”autistic” behaviour when instantiated in the predictive processing framework.

In this section, we implemented and trained predictive processing models to discriminate between MNIST digits. We trained an 'autistic' and a neurotypical network. The autistic network had about ten percent of it's connections between layers zeroed out to simulate the poorer long distance connectivity thought to exist in autism. The layers of the neurotypical network were all fully connected with no weight mask applied. We showed that although the autistic networks performed slightly poorer during training than the neurotypical networks, they performed better at the validation task. Moreover, the autistic network was significantly more sensitive to identity-preserving rotation and translation transformations of the images than the neurotypical network was, which confirms our main hypothesis. We have thus constructed a predictive processing neural network model which can mimic some aspects of autistic functioning, and this therefore provides empirical support for the under-connectivity hypothesis as well as our predictive processing account of autism

3.4 Integrative networks

3.4.1 Methodology

The second aspect of our hypothesis is integration. We argue that, due to the impaired long range connectivity, autistic brains have a lesser ability to integrate information across multiple modalities. We set out to test this assumption via a predictive processing model. This neural network model had to integrate two different modalities of

information - colour and MNIST digits. A simple diagram of the network architecture is shown below:

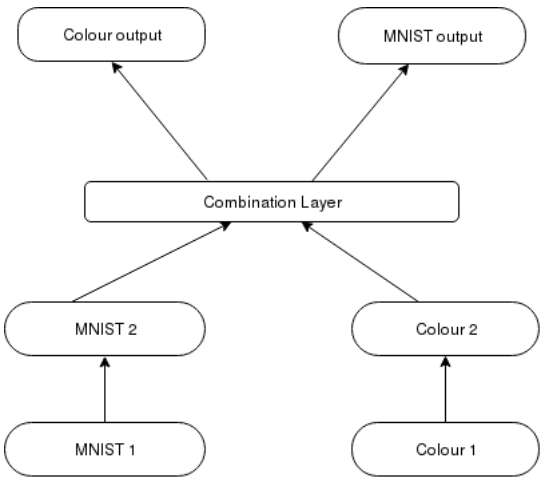

Figure 3.8: A simple diagram of the network architecture of the intregrator network. Each modality had its own two separate layers of base network, simulating the lower sensory cortices. Their information was then combined into a combination layer which then had to predict the colour and digit again. The combination layer represented a higher, multimodal, sensory region.

Each of the separate modalities - the colour and the MNIST digit - had it's own base network for the first two layers. These networks simulated the sensory cortices. The activations in these layers was then combined into a single combination layer which took inputs from both base networks simultaneously. This layer represents an integratory higher level region of the brain - as both stimuli are visual in nature this would represent the high level visual area IT.

The combination layer must then integrate its two sets of sensory inputs to predict the form of its inputs. We adapted this to a classification task as follows: the Euclidian distance between the predicted output and the actual output was calculated. If this distance was smaller than that between the output and any other class, then the network was said to have classified the inputs correctly. The prediction error was calculated also as the Euclidian distance between the predicted and desired output, and these prediction errors were backpropagated through the system.

The deficit in long range over short range connectivity was achieved by applying a

mask over the weight matrices connecting the layers of the autistic network. This mask zeroed out a certain degree of connections randomly. Since we did not suspect that the connectivity impairment in autism is particularly drastic, we chose a value such that approximately 10% of all connections in the autistic network were zeroed out. As we predict that integration in closely coupled sensory regions is not as impaired as that in longer range connections, we applied no masking matrices to the lower two layers of the network. To simulate local overconnectivity in sensory regions of the networks with autism, we increased the number of hidden units in the lower regions from 500 to 1000 in the networks with autism.

3.4.2 Results

Our first hypothesis is that due to poorer long range connectivity, the autistic networks will perform worse than controls at the task, which requires a successful integration of both stimuli.

The training plots of the two networks are shown below:

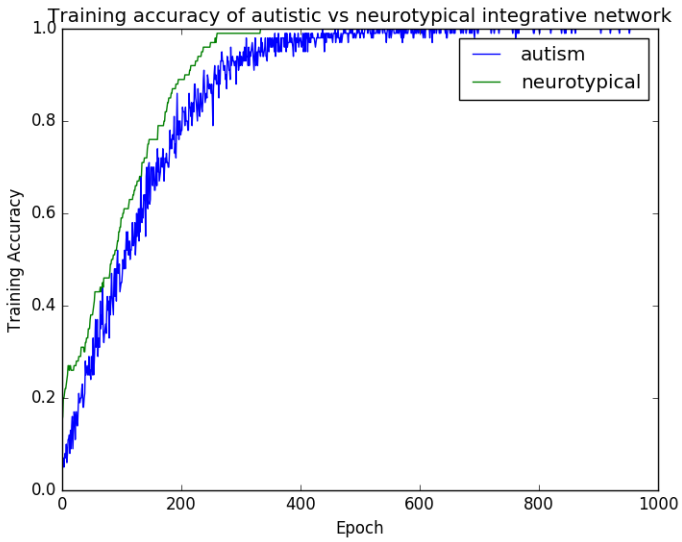

Figure 3.9: The accuracy of the integratory network during training for both the autistic and neurotypical network

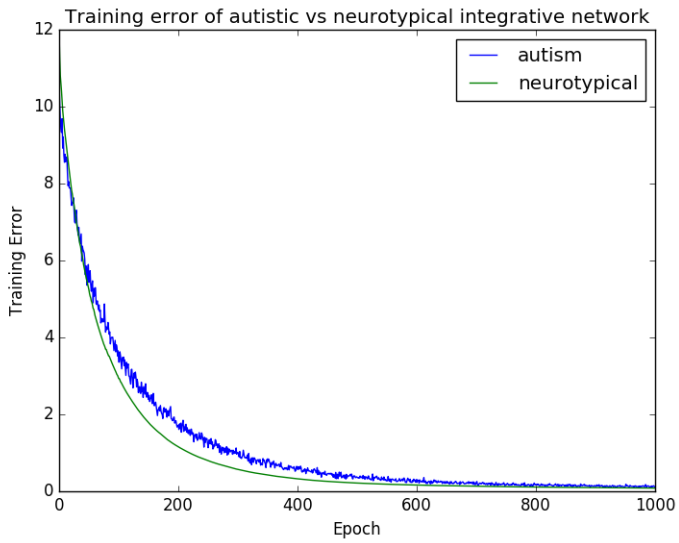

Figure 3.10: The error of the integratory network during training for both the autistic and neurotypical network

Empirically the autistic networks do train slower, and there is significantly more variance in the training path of the network with autism. We hypothesise that this might be a result of the under-connectivity between high and low areas which may impair the correct transmission of predictions. This would have the effect that the predictions as they are experienced by the lower levels have significantly more variance than the ”true” predictions would, thus leading to greater variation during training as the lower levels adjust to try to match the corrupted predictions.

A plot of the validation accuracy overall obtained from both the autistic and control network is plotted below:

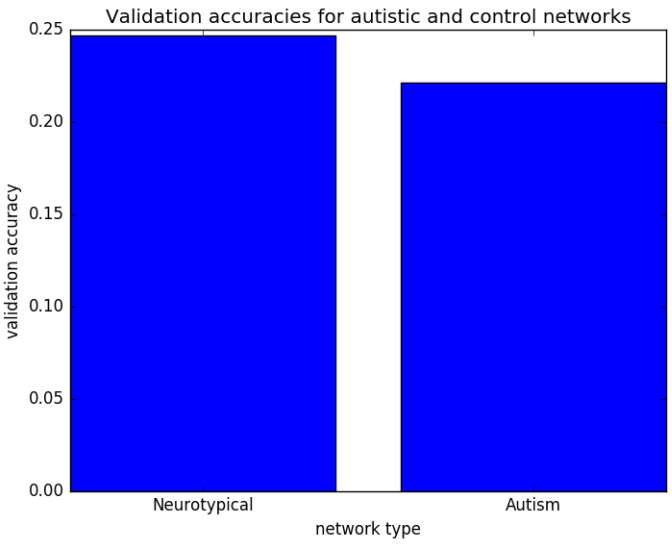

Figure 3.11: The validation accuracy of the integratory network for both the autistic and neurotypical network

It appears that, even on the validation data, the autistic network performs slightly worse than the control, thus validating the hypothesis that the pattern of impaired long range connectivity between regions can negatively affect the ability of the network to integrate multiple sources of information together effectively.

A further prediction generated from our predictive processing model is that of a decreased precision value for the prior predictions in autism. This is because the predictions propagating down the hierarchy are impaired due to the poor long-range connectivity. As the predictions are less accurate and useful in predicting the sequence of inputs flowing up to that layer of the hierarchy, then each layer will naturally downweight the importance of the predictions, thus causing the ”precision” of the predictions to decrease. This provides a solid neurophysiological mechanism for the decrease in precision as postulated by Lawson et al 2014 [71] as an explanation for autism. To test this hypothesis we added precision variables to the model which were operationalised as weights on the cost function to determine the relative importance of the prior predictions or the sensory evidence. These weights, being simple scalar variables, and interacting with the rest of the cost function by simple elementwise multiplication, are differentiable, and thus themselves can be optimised by gradient descent. The evolving precision weights during training are plotted below for both the autistic and neurotypical network:

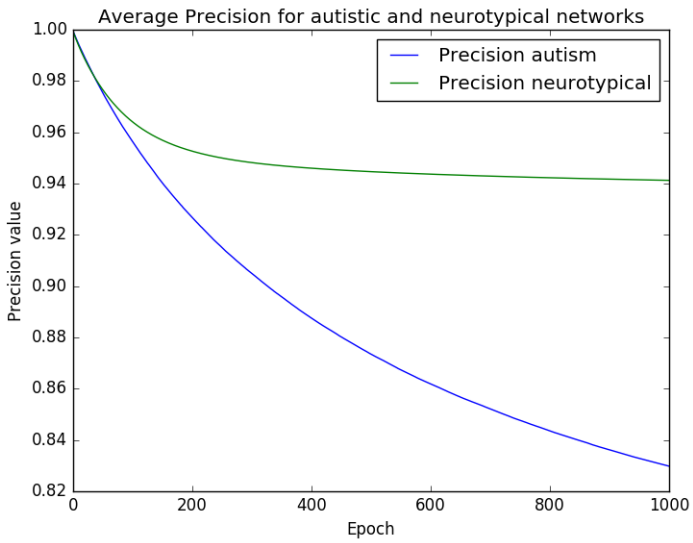

Figure 3.12: The evolving value of the precision weights for both neurotypical and autistic networks over the course of training

The graph shows that although both precisions initially decrease, the precision of the neurotypical network rapidly plateaus while that of the autistic network continues to decrease. This is exactly as predicted by the hypothesis which states that we expect the precision of the predictions in autism to be substantially lower in neurotypical networks. This thus provides a solid vindication of the modelling approach pursued here as the approach can not only replicate some of the external behaviour seen in autism, but also some of the internal mechanisms independently proposed to account for such behaviour.

In this section, we implemented and trained an integrative predictive processing neural network which had to integrate information from two different modalities - greyscale image and colour - in order to complete its task. The autistic network, as before, had a mask applied to its weight matrices which zeroed out a fraction of the connections between the combination layer and the lower level sensory layers. For the neurotypical network no mask was applied. The results show that, as predicted by the hypothesis, the autistic network is poorer at the integratory task in both the training and validation phases. This agrees with a wide body of empirical evidence showing that individuals with ASD often perform worse at psychophysical tasks requiring the integration of information from different modalities, and thus our model can be successfully used to capture a facet of an autistic deficit. Moreover, when precision variables were added to

the model, they followed the pattern predicted both by our novel predictive processing hypothesis of autism but also that of Lawson et al (2012).

3.5 Autistic Savantism

3.5.1 Methodology

There are three main hypotheses relating to the hemispheric model of autism. The first is that the autistic models with low hemispheric connectivity should have greater variance in performance. The second is that the mean of the autistic models should be slightly lower than control models, and the third is that some networks should exhibit savant-like behaviour due to the more autonomous specialisation of the hemispheres.

We designed a network architecture to test these hypotheses. The network was set up such that there was a split between the hemispheres. Each hemisphere was given half of the input and from that half, had to predict what the other half of the input was. Thus, the broad setup of the network was similar to that of the split-brain autoencoder of Zhang et al 2016 [102]. However we also adjusted the degree of interhemispheric connectivity. This connectivity was implemented through recurrent connections between the hemispheres. The degree of connectivity was operationalised by the number of allowed connections.

The network was composed of a single hidden layer projections to the input and output layers, as well as recurrent connections to the other hemisphere. The default parameters for this network were 16 hidden units, a learning rate of 0.005, and training with gradient descent using the Adagrad optimser [103]. To obtain robust statistical information about the effect of the hemispheric split, a population of 100 networks were trained in each experiment. Each network was trained for 500 epochs.

A diagram of the network architecture is shown below:

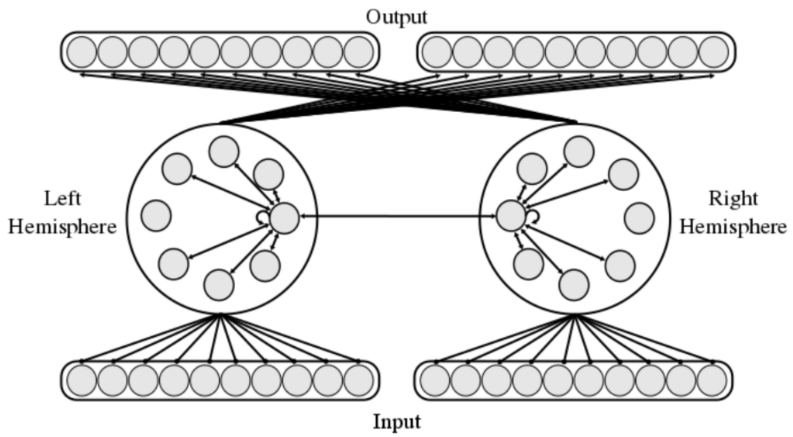

Figure 3.13: A diagram of the network architecture. The network is split into two halves, or hemispheres, with some degree of connection between them. This connectivity will be the crucial hyperparameter investigated.

The inputs to the network were abstract bit patterns, each 20 bits long. These were arranged into skills, in which the patterns for each skill were generated in a systematic way, thus producing a statistical grouping where patterns within one skill were generally much more similar to each other than patterns outside of the skill. For instance one set of skills used correlated bit patterns - each skill corresponded to a set of patterns generated with a specific correlation coefficient.

To ensure valid statistical generalisation of our results, instead of merely training a single network, a population of 100 networks was trained and evaluated for all tests. Each network received a unique uniform initialisation of weights according to a Gaussian distribution centred at 0 with a standard deviation of 0.05. Apart from their random initialisation, all the networks in the population were identical.

The hypothesis is that having impaired connectivity between hemispheres allows each hemisphere to specialise more than each could otherwise. We expect this effect to manifest in a number of ways. The first is the variance of output of the networks with impaired connectivity should be greater. This is because, since each hemisphere is independent, their contributions to the final output are more independent, thus leading to more variance by the simple mathematical argument that the variance of the composition of two independent variables is larger than that of two correlated variables.

The second is that we expect the autistic networks to perform slightly worse, on average, than the control networks. Conversely we expect that the autistic networks to perhaps show some evidence of savant-like behaviour - significantly better performance on one skill than the average. This is because the greater specialisation allowed by the impaired interhemispheric connectivity can lead to a hemisphere specialising in a single skill, while this specialisation must cause the other skills to be degraded in relative terms since the number of neurons and connections, and thus the representational capacity of each hemisphere, is limited.

3.5.2 Results

For each skill 100 networks were trained for 100 epochs. The number of inter-hemispheric connections in each network were varied from only 1 connection to a full 7 connections. As the number of inter-hemispheric connections was varied, the number of intra-hemispheric connections was kept constant at 6. The training curves of each inter network averaged over all skills and all networks were plotted. These are shown below:

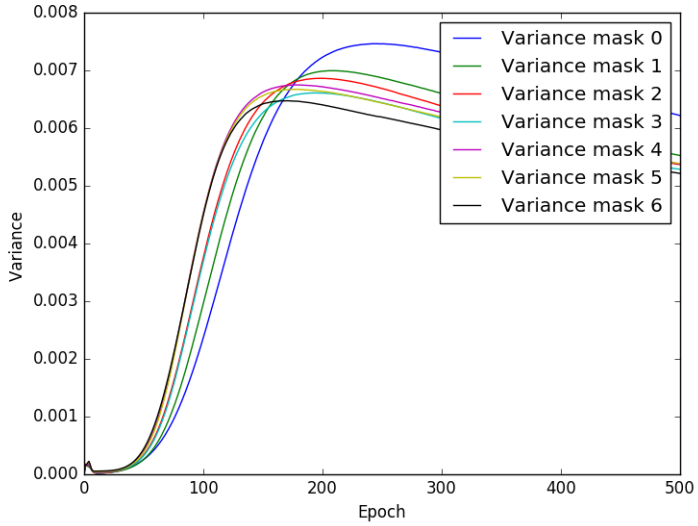

Figure 3.14: The variance of the training curves averaged over all networks and skills for each degree of interhemispheric connectivity. The intrahemispheric connectivity was 5.

Although the variance rises dramatically and then declines over the course of training, a clear correlation can be observed between the number of interconnections and the variance of the output network towards the end of the training scheme. This agrees with the hypothesis which states that networks with fewer cross-connections, and therefore lesser connectivity to have increased variance compared to networks with significantly greater cross connectivity.

The second hypothesis is that networks with impaired connectivity between hemispheres are expected to perform worse at the task in general than the control networks. This is because the recurrent cross-hemispheric connections allow them to utilise information seen by the other hemisphere. A bar chart of the averaged MSE error across networks and epochs for each skill is shown below:

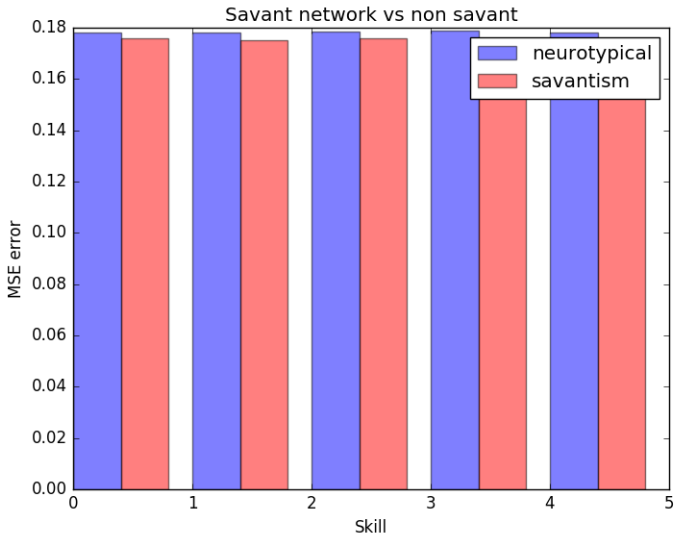

Figure 3.15: The average MSE achieved by an average of the population of networks across each skill. The ”autistic” network is that with 2 cross-connected units. The ”neurotypical” network has 6

From this we observe that our hypothesis is falsified. The autistic networks actually appear to perform better at the task, as measured in MSE error, than the non-autistic networks. We hypothesise that this is due to the fact that the skills were generated in such a way that there is no temporal dependence between each skill in the batch, while the cross-hemispheric connections were recurrent, meaning that they only applied to the network in the next time step. Since there were no effective temporal dependencies between each successive input to the network, this would have the effect of simply

injecting noise into the network in proportion to the degree of cross-connectivity, thus impairing performance. We tested this hypothesis by designing temporally successive skills in which each bit of the input was correlated with the same bit in the previous input. This meant that the recurrent interhemispheric connections could convey useful information about the skill at the next timestep. The averaged MSE error of an ”autistic” network with 2 cross-connections and a neurotypical network with 6 cross-connections, averaged across the final 100 epochs and all the network in the population are presented below:

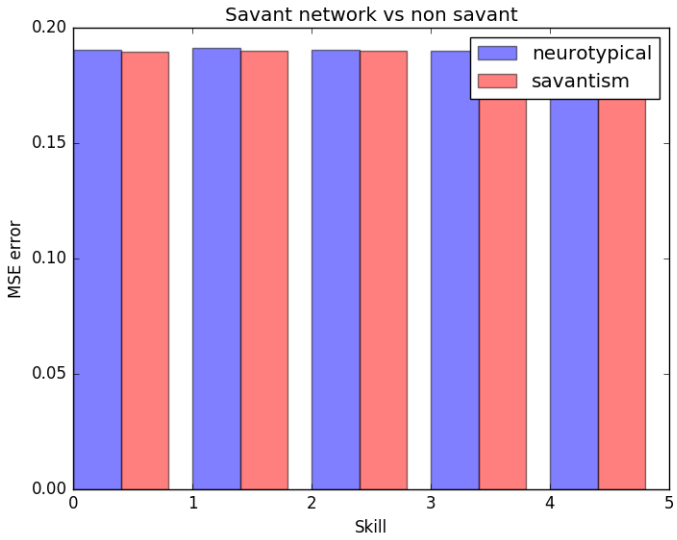

Figure 3.16: The average MSE achieved by an average of the population of networks across each skill. The ”autistic” network is that with 2 cross-connected units. The ”neurotypical” network has 6

These results show that with when these skills which include temporally sequential information are tested, the average of the autistic and the neurotypical networks are approximately identical. Thus it appears that using data containing sequential information has eliminated the superior performance of the autistic networks. However, it is interesting that despite needing some degree of interhemispheric connectivity to predict the data, the autistic networks still perform approximately equal to the neurotypical networks which contained many more cross-connections. One possibility is that the two cross-connections in the autistic network proved sufficient for the transfer of enough temporal information to allow the development of a good model of the input data. We tested this hypothesis by comparing the control network as above with a network with only a single cross-connection between the hemispheres.

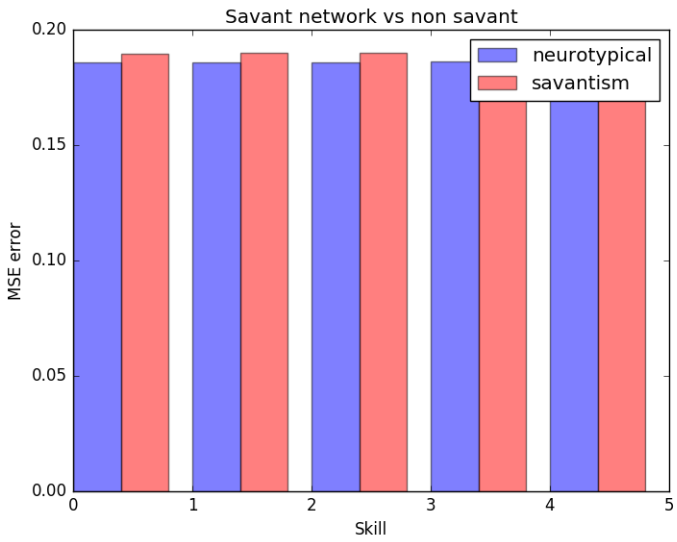

Figure 3.17: The average MSE achieved by an average of the population of networks across each skill. The ”autistic” network is that with 2 cross-connected units. The ”neurotypical” network has 6

With only a single cross-connection, the autistic networks perform slightly worse than the neurotypical networks, implying that only a single cross connection is insufficient to convey enough information about the stimulus across time steps to use to predict the data accurately. When a network with no cross connections is used, performance is significantly worse, as expected.

It is also important to note that although is no evidence of the savant-like skills in the averaged analysis, this is unsurprising since we expect the savant skills to be somewhat rare and also distributed randomly among the five skills, meaning that any fluctuations there are averaged out in these plots.

To obtain evidence for savant-like skills, the results were broken down still further to plot the MSE obtained by a single network in the population for a given setting of inter and intra hemispheric connectivity. For some networks a savant-like pattern of skill is detected. For instance, the chart for network 76 for epoch 500 is plotted below:

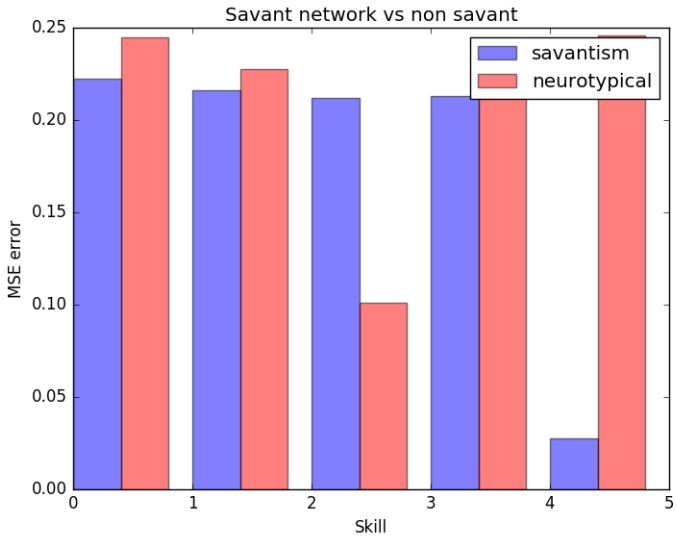

Figure 3.18: A graph of a network which has developed a savant skill which is significantly better (in terms of lesser MSE errors) than the other skills.

The autistic network possessed 1 connection between each hemisphere while the neurotypical network possessed 6. As can be seen, there is some degree of savantlike skill where one skill in the autistic network is significantly better than the rest. However, it is important to note that the general error for the autistic networks, even for the non-savant skills, was often lower than that of controls. This is likely due to the non-sequential nature of the inputs, as explained above. Savant-like patterns are not robust, however. Many networks do not show any evidence of savant-like skills. This pattern is shown in the graph below:

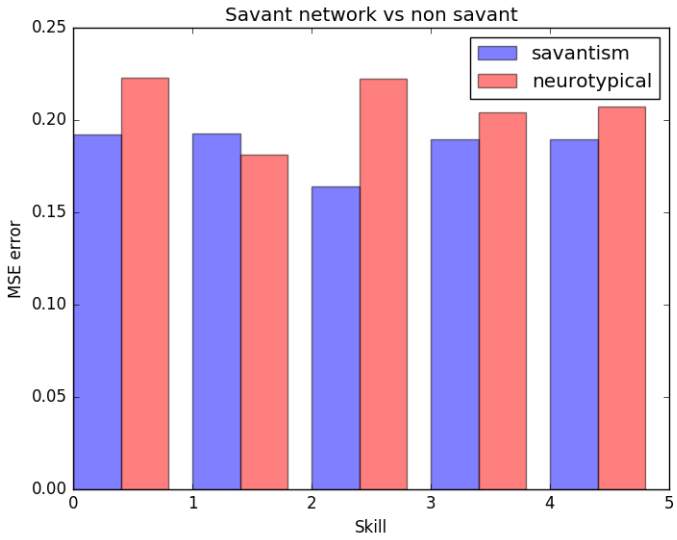

Figure 3.19: A graph of a network which does not show any apparent savant-skill. This shows the MSE error for network 89 in the final 500th epoch

This is not surprising, as savant skills are uncommon even in populations diagnosed with ASD. The precise mechanisms that drive some networks to develop a savant skills while others do not are difficult to determine.
In addition, in some cases, both the autistic and the neurotypical network exhibit what appear to be savant skills, as is shown in the plot below:

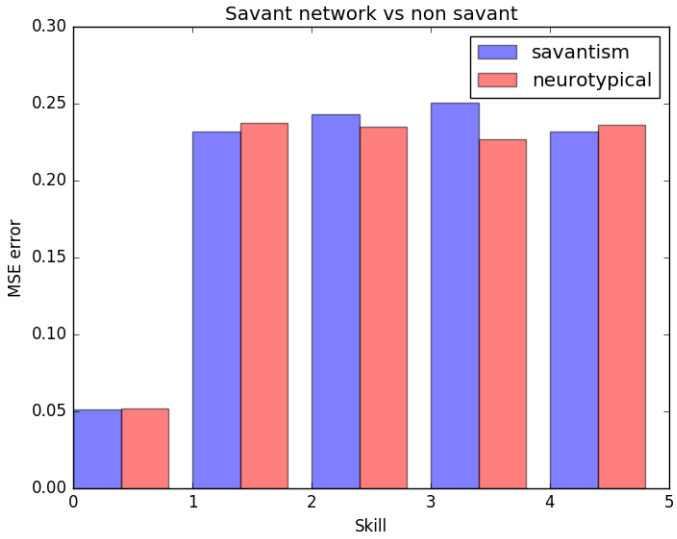

Figure 3.20: A graph of a network in which both the neurotypical and the autistic networks appear to display savant-like skills.

Exactly why this occurs is unclear. One possibility is that the degree of connectivity even in the supposed neurotypical networks is insufficient to prevent hemispheric specialisation from developing, and thus all networks tested here are all ”autistic” in a sense. Another is that the inherent nature of the task and network architecture, means that the network tends towards savant-like learning, no matter the degree of inherent connectivity. A third possibility is that this pattern is due to a defect in the training procedure. It is possible, for instance, that the ”savant” skill is simply that which the network learnt first or last, for instance, and therefore has no relation to the degree of connectivity. Under this hypothesis, it is difficult to explain, however, why many networks should show no evidence of savant skills.

One possible confounder of results is that the total number of connections differed between networks since the number of inter-hemispheric connections was varied while the number of intrahemispheric connections remained the same. It is thus possible that the results observed are simply an artefact of the that networks with greater interhemispheric connectivity simply have greater total connectivity, and hence greater representational power. The greater lesser variance in the performance of the networks with greater interhemispheric connectivity may simply because they better train to match the input stimuli and a greater capacity overall so that perhaps their success is less dependent upon a certain training path. There appears to be little evidence for this, however, since the mean errors of the greater and lesser interhemispherically connected networks were so similar, and also this hypothesis would struggle to explain the development of savantism in any network. Nevertheless, this hypothesis was tested by creating networks which varied the number of intra-hemispheric connections in a manner inversely proportional to the number of interhemispheric connections, so that the total number of connections remained the same at all time. The variance over 2000 training epochs for all values of interhemispheric connectivity - from 1 to 7 - was plotted. The number of total connections in the network was held total at 8, and thus a network with seven interhemispheric connections possessed one intrahemispheric connection, and vice versa.

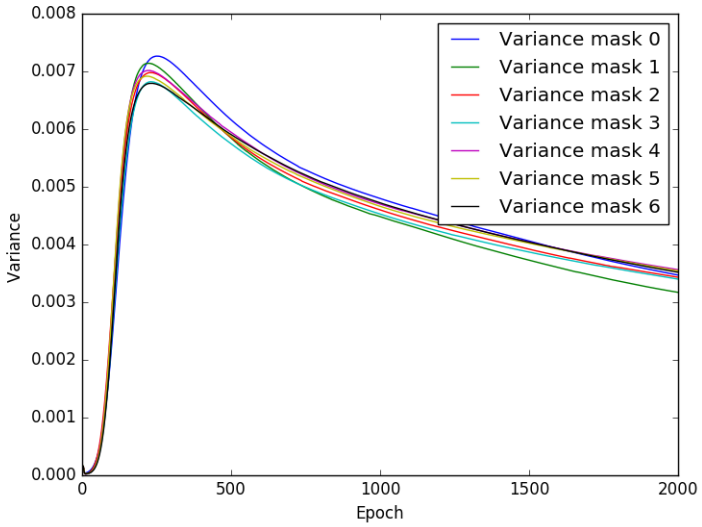

Figure 3.21: The training variance over time of networks with an equal number of total connections.

This graph shows a similar pattern to the others. Thus it appears that the total number of connections in the network is not a significant confounder of the results.

In this section, we have implemented and tested the effect of the degree of network hemisphericity in autism, and in the development of autistic savantism. We have shown that by partitioning a network into two hemispheres, the variance in performance of the population of networks is greater when there is a smaller amount of interhemispheric connectivity. We have also shown that in at least some of the networks, savant-like skills can develop in that the network shows a pattern of performance where it performs significantly better at one skill than the others. We show that savant skills do not develop in all networks, however, which suggests that they are not produced purely as an artefact of the training process or network architecture. We also show that the results obtained are not simply a result of there being a greater number of connections in total in the model, but are actually due to the differences in interhemispheric connectivity.

Chapter 4

Discussion

Our predictive processing hypothesis of autism proposed three key hypotheses which we set out to test. These were that the pattern of long-range underconnectivity and short range overconnectivity observed in studies of autism would lead several seemingly disparate effects: first, it would cause the lower levels of the sensory hierarchy overfitting the data, thus leading to improved discrimination on some tasks, but also impaired generalisation and increased sensitivity to small and irrelevant differences in stimuli compared to controls. Second, due to the impaired long range connectivity, the predictions flowing down the hierarchy to the lower levels are incomplete, thus helping contribute to the overfitting, and also making the lower levels less sensitive to high level context. Third that the impaired connectivity between layers of the hierarchy would mean that information about the sensory stimuli would have difficulty being propagated up the hierarchy, thus causing the higher levels to generally become impoverished relative to controls and leading to worse overall performance at tasks that require complex analysis of more than local features, and also a difficulty integrating multiple disparate sources of information.

We tested the first hypothesis - that lower levels of the sensory hierarchy overfit - by training a discriminative network to discriminate between different mnist digits, and then between small variations on the same mnist digit. For both translations and rotations, the autistic networks with poorer connectivity showed a greater sensitivity to these identity-preserving transformations than controls, thus vindicating this part of the hypothesis. On the other hand, it was found that despite this greater fine-grained sensitivity, the autistic networks actually performed slightly worse at the simpler task of learning to discriminate the two digits. One possibility here is that this worse over-

all performance could be due to the greater sensitivity to small, irrelevant differences, thus leading to the autistic networks to reject a pair of the same digit due to slight differences between them. This kind of brittle cognition is also a feature of autism. On the other hand, few studies report a generalised perceptual impairment in autism, and thus this aspect of the result should not be taken as representative of autism generally. Moreover, the advantage of the neurotypical network over the autistic one in the simple MNIST discrimination task was fairly small compared to the significantly greater sensitivity of the autistic network to the slightly transformed digits,

The second hypothesis was that due to the poorer long-range connectivity, the information reaching the higher areas is disrupted. This is especially the case when the higher areas must integrate information from multiple regions and modalities together. Because of this we hypothesise that a network which must integrate information from two different sensory regions together will perform poorer when instantiated with the autistic pattern of connectivity than when it has a more standard neurotypical pattern. We found that the autistic networks did indeed train slower, plateaued at a higher level of error and obtained worse accuracy on the training set than the control networks. They also performed worse in validation testing, thus confirming that the effect is not merely due to overfitting on the training set. Moreover, the training path of the autistic networks had significantly more variance than control networks. One possible reason for this is that, due to the impaired connectivity between layers, the predictions propagated down from the higher level were distorted and disrupted, thus leading to incorrect predictions reaching the lower levels, thus causing the gradient descent learner based on the prediction errors to update wrongly in some instances, thus leading to greater variance in training. Both of these results were in line with the initial hypothesis, and this provides strong evidence that a connectivity pattern similar to that observed in autism can cause significant deficits in the abilities of networks to integrate multiple disparate stimuli together to respond to a task.

Additionally we tested the theory of Lawson et al (2014) that the precision given to the prior predictions in autism is lower than in neurotypicals, thus leading to individuals with ASD tending to overweight the importance of local sensory information over more global higher level gestalts. While Lawson et al did not provide any kind of mechanistic explanation for why the precision should be downweighted in autism, except perhaps due to a imbalance of neurotransmitters, our theory provides the ex-

planation that due to the impaired connectivity, the predictions flowing down will be disrupted and noisy, and thus worse at predicting the inputs flowing up the hierarchy than in neurotypicals, and that thus each layer has a strong incentive to downweight their precision. We tested our hypothesis by adding prediction variables to the model and found that in the networks with the autistic pattern of connectivity, they naturally decreased unlike those with a neurotypical pattern, thus confirming the hypothesis. However, in our model the precision variables are optimised by gradient descent which is the same process that updates the weights. This is different to most predictive processing models which tend to assume the precision is a more global state determined by general concentrations of various neurotransmitters, and adapts or updates in a different way to the synapses which carry the equivalent of the 'weights' of the network. Because of this mismatch, it is unclear to what extent our results can be generalised to the biological setting. Nevertheless, our model provides at least an existence proof of the mechanism of decreased precision presented in our predictive processing theory of autism.

A related hypothesis concerns the pattern of connectivity in autism and the development of autistic savantism. This argument relies on the connectivity of the hemispheres of the brain. The corpus callosum appears to be smaller in autism as compared to neurotypical brains and this is consistent with the long range pattern of long-range underconnectivity and short range overconnectivity. It is hypothesised that because of the lesser interhemispheric connectivity, each hemisphere can learn and develop skills more independently than in a fully connected brain. This, in turn, means that hemispheres can specialise for individual skills in those with the autistic pattern of connectivity than neurotypicals, and generally are much less coordinated in developing and learning specific capacities. This increased specialisation can yield especially good results if the specialisations learned by each hemisphere complement each other, but, conversely, will yield especially bad results if they do not. This means that it is expected that the performance of populations of networks on various tasks will have greater variance than control networks. Moreover, this greater autonomy and variance of hemispheres could lead to one hemisphere specialising in one skill to the expense of the other, thus leading to the development of savant skills. It is worth noting that although this argument is couched in terms of hemispheres, it can be relatively easily extended and generalised to talk about relations between any functionally separate units in the brain

A network with a simulated hemispheric split was implemented to test these hypotheses. It was found that, in accordance with the first hypothesis, the variance of training error across the population of networks decreased monotonically with the amount of interhemispheric connectivity, thus confirming that hypothesis. It was, however, observed that in general the networks with the pattern of connectivity prevalent in ASD often performed slightly better on average than the control networks. Exactly why this is so remains unclear, since it is naively expected that the autistic networks perform worse, or at about the same level as controls. One possibility is that it is due to the fact that the nature of the task and networks were such that greater hemispheric autonomy and specialisation was generally beneficial in this task as compared to the multitude of more complex tasks the brain faces in the real world. Since our data were so primitive-just correlated streams of bits - this question of ecological validity is of real importance.

Some individual networks in the population were observed to possess something like a savant skill - a single skill on which the network performed significantly better than all the other skills. Savant networks were not ubiquitous, however they were quite common - significantly more than the proportion of savants in the population with ASD. Although, our model, being merely an existence proof of the importance of hemispheric specialisation and connectivity playing an important role in the development of some facets of autistic behaviour, specifically savant skills, it should not necessarily be expected to reproduce the exact proportions of savantism found in the population with ASD. The increased proportion is intriguing, however. This is especially the case when it is considered that on occasion the supposedly neurotypical networks also exhibit cases of savant skills. One possibility is that the nature of our autoencoding task or the network architecture naturally predisposes it towards specialising in one skill at the expense of the others. This may be because, since the skills are only patterns of random bits with different degrees of correlation, and although the skills are presented in sequence, they are each presented to the same input layer with the same weight matrix. This means that the network must use the same pattern of connectivity with every skill, which renders with the choice of either specialising in one skill to the detriment of the others, or else learning a weight configuration which is poor for all skills, but nevertheless better than the configuration specialised at only one skill. This may explain why many even supposedly neurotypical networks developed savant-like skills in this model. However, it is interesting that different networks in the population devel-

oped different strategies, as it shows that the particularities of the random initialisation is enough to have a very large effect on the later configuration of the network.

Overall, however, the results show that this pattern of connectivity between hemispheres does leader to greater variance in the result, and furthermore can provide a mechanism by which savant skills can develop. Moreover, even though some of the neurotypical networks developed savant skills unexpectedly, the proportion of savants was greater in the autistic networks with lesser cross-hemispheric connectivity, thus confirming the hypothesis at least to some degree.

Moreover, the results of the experiments with the predictive processing model broadly conform to the predictions of the theory of predictive processing in autism proposed in this paper. These results, then, provide an existence proof, and some empirical evidence in support of the theory, even though the models used are, by necessity, significantly high-level and abstracted away from the biological detail.

4.0.1 Limitations

All of the networks in this paper used artificial neural networks to model biological systems. However, it is unclear the extent to which the dynamics observed in artificial neural networks may generalise to more complex biological systems. Unlike a McCullough and Pitts neuron, biological neurons cannot simply be understood as weighted sums of their inputs; instead many factors other than the direct inputs are involved in the firing mechanism. Such firing, moreover, cannot simply be understood as a single scalar value, but instead is a temporally extended spike train with complex internal dynamics which may convey additional information about the signal. Moreover, weights and firings in artificial neural network models can take any scalar value while real neurons are constrained to a maximum firing rate, and also are typically either exhibitory and inhibitory, but not both. The more complex nature of biological systems means that a biologically plausible model of a similar system might give rise to different dynamics to those observed in the current models. The networks were also trained by gradient descent using backpropagation of error, a method which is assumed not to be biologically plausible, although there have been attempts made to justify backpropagation in a biological setting [104], or else to create new algorithms which can

implement an approximation of backpropagation in a biologically plausible way [105].

Additionally, the networks used to model the phenomena were relatively small, and only contained several distinct layers at most which were arranged into a simple hierarchical configuration. The brain, by contrast, is composed of multitudes of independent regions, each connected to a complex web of other regions instead of a clear hierarchy. Due to this difference, it is unclear whether the dynamics observed by altering the connectivity of the simple, hierarchical, artificial models can be fruitfully generalised to the complex, heterarchical brain. This is especially the case in the integratory network which only had to integrate inputs from two different regions together. Most regions of the brain must integrate information from a substantially larger number of regions than two, and so the model of only two different modalities may prove insufficient for modelling the challenges of regions which must integrate many more. On the other hand, the model does, at least, provide an existence proof that in a simplified scenario, the predictive processing hypothesis is supported.

These considerations of biological plausibility and model scale mean that evidence derived from our models should not be considered to apply to the brain in current form. Instead, it should be considered as validation that an abstract informational system, when restricted to the patterns of connectivity and communication observed in studies of autistic individuals, exhibits abstract analogues of many of the observed deficits in autism, and explains why several seemingly disparate impairments are so often clustered together.

Furthermore, due to computational tractability concerns, as well as the greater literature and knowledge dealing with artificial neural networks over fully probabilistic Bayesian systems, our predictive processing model did not utilise fully Bayesian computation and fully specified distributions, but instead approximated the distribution by a single point or layer of activations. Although it still remains unknown whether the brain also uses full distributions in its reasoning, or else some approximate form, our use significantly diverges from the guidelines laid down by Friston et al, which may limit the applicability of the model.

4.0.2 Future Work

The models presented in this paper were often fairly simple and intended to provide an abstract proof of concept for the concepts and theories discussed. In addition, they are not particularly biologically plausible. As such the results and evidence derived from them cannot be directly applied to the brain. For subsequent research it could be interesting to make the networks described in this paper more biologically plausible, for instance by utilising more biologically plausible learning rules such as Hebbian plasticity or Spike Time Dependent Plasticity (STDP). Moreover, more complicated network architectures, modelled directly on specific regions of the brain could be experimented with, as well as using more ecologically valid stimuli and datasets to test and train the networks on. An example of this could be to train the model of autistic savantism not simply on random correlated bit patterns, but on ecologically valid skills that the brain needs for survival, such as object recognition, segmentation and so forth. This would naturally require a considerable scaling up of the model, but the results would also be significantly more compelling.

Additionally, relatively little exploration of the parameter space was undertaken. Although this means that the sensitivity of the results to the configuration of hyperparameters cannot be too precise, since the default hyperparameters resulted in the evidence presented, no quantitative measure of the sensitivity of the results to the hyperparameters is known. Nor is how the results obtained depend upon the features of the network architecture such as the number of layers or the how the inputs are combined in the integratory network. Future work could significantly improve our understanding of the effects of hyperparameter and architectural choices on the performance and behaviour of predictive processing models.

Overall, the successes of the model in validating the novel predictive processing hypothesis of autism suggest the productivity of the approach of integrating the insights from the neurophysiological and theoretical levels of neuroscience to produce novel theories of brain disorders such as autism.

4.1 Conclusion

In this dissertation, we have proposed a novel predictive processing explanation of autism which integrates insight from both the neuroscientific and theoretical level, we have also provided neural network models that demonstrate that the patterns of aberrant connectivity found in individuals with ASD, when instantiated in a predictive processing model, can organically give rise to several seemingly disparate deficits that characterise autism. Moreover, we also applied the same paradigm of abberant connectivity to interhemispheric connectivity in autism and showed that such a pattern of connectivity between hemispheres can lead to greater variance in the autistic network as well as possibly the development of savant skills in some of the networks, thus providing a mechanistic model of a fairly mysterious phenomenon.

Bibliography

[1] Simon Baron-Cohen, Alan M Leslie, and Uta Frith. Does the autistic child have a theory of mind? *Cognition*, 21(1):37–46, 1985.

[2] Uta Frith and Francesca Happé. Autism: beyond theory of mind. *Cognition*, 50(1):115–132, 1994.

[3] Karl Friston. A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1456):815–836, 2005.

[4] Jon Brock, Caroline C Brown, Jill Boucher, and Gina Rippon. The temporal binding deficit hypothesis of autism. *Development and psychopathology*, 14(2):209–224, 2002.

[5] Michelle Turner. Annotation: Repetitive behaviour in autism: A review of psychological research. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(6):839–849, 1999.

[6] Francesca Happé and Uta Frith. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36(1):5–25, 2006.

[7] Francesca Happé, Ronald Angelica, and Robert Plomin. Time to give up on a single explanation for autism. *Nature neuroscience*, 9(10):1218, 2006.

[8] Francesca GE Happé. Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. a research note. *Journal of Child Psychology and Psychiatry*, 37(7):873–877, 1996.

[9] Danielle Ropar and Peter Mitchell. Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *Journal of child Psychology and Psychiatry*, 42(4):539–549, 2001.

[10] Leo Kanner et al. Autistic disturbances of affective contact. *Nervous child*, 2(3):217–250, 1943.

[11] Hans Asperger and Uta Trans Frith. 'autistic psychopathy'in childhood. *American Psychological Association*, 1991.

[12] Mayada Elsabbagh, Gauri Divan, Yun-Joo Koh, Young Shin Kim, Shuaib Kauchali, Carlos Marcín, Cecilia Montiel-Nava, Vikram Patel, Cristiane S Paula, Chongying Wang, et al. Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, 5(3):160–179, 2012.

[13] Craig J Newschaffer, Lisa A Croen, Julie Daniels, Ellen Giarelli, Judith K Grether, Susan E Levy, David S Mandell, Lisa A Miller, Jennifer Pinto-Martin, Judy Reaven, et al. The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health*, 28:235–258, 2007.

[14] Michael Rutter. Incidence of autism spectrum disorders: changes over time and their meaning. *Acta paediatrica*, 94(1):2–15, 2005.

[15] Daniel H Geschwind. Genetics of autism spectrum disorders. *Trends in cognitive sciences*, 15(9):409–416, 2011.

[16] Joachim Hallmayer, Sue Cleveland, Andrea Torres, Jennifer Phillips, Brianne Cohen, Tiffany Torigoe, Janet Miller, Angie Fedele, Jack Collins, Karen Smith, et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of general psychiatry*, 68(11):1095–1102, 2011.

[17] Joseph Piven, Pat Palmer, Dinah Jacobi, Debra Childress, and Stephan Arndt. Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry*, 154(2):185–190, 1997.

[18] Elena Bacchelli and Elena Maestrini. Autism spectrum disorders: molecular genetic advances. In *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, volume 142, pages 13–23. Wiley Online Library, 2006.

[19] Nicolas Ramoz, Jennifer G Reichert, Thomas E Corwin, Christopher J Smith, Jeremy M Silverman, Eric Hollander, and Joseph D Buxbaum. Lack of evidence for association of the serotonin transporter gene slc6a4 with autism. *Biological psychiatry*, 60(2):186–191, 2006.

[20] Tero Ylisaukko-oja, Karola Rehnström, Mari Auranen, Raija Vanhala, Reija Alen, Elli Kempas, Pekka Ellonen, Joni A Turunen, Ismo Makkonen, Raili Riikonen, et al. Analysis of four neuroligin genes as candidates for autism. *European journal of human genetics: EJHG*, 13(12):1285, 2005.

[21] Charles A Williams, Aditi Dagli, and Agatino Battaglia. Genetic disorders associated with macrocephaly. *American journal of medical genetics Part A*, 146(15):2023–2037, 2008.

[22] Raymond J Kelleher and Mark F Bear. The autistic neuron: troubled translation? *Cell*, 135(3):401–406, 2008.

[23] Damon T Page, Orsolya J Kuti, Chrysa Prestia, and Mriganka Sur. Haploinsufficiency for pten and serotonin transporter cooperatively influences brain size and social behavior. *Proceedings of the National Academy of Sciences*, 106(6):1989–1994, 2009.

[24] Thomas Bourgeron. A synaptic trek to autism. *Current opinion in neurobiology*, 19(2):231–234, 2009.

[25] Mark S Boguski and Allan R Jones. Neurogenomics: at the intersection of neurobiology and genome sciences. *Nature neuroscience*, 7(5):429, 2004.

[26] Alan M Leslie and Francesca Happé. Autism and ostensive communication: The relevance of metarepresentation. *Development and psychopathology*, 1(3):205–212, 1989.

[27] David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526, 1978.

[28] Alan M Leslie and Uta Frith. Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, 6(4):315–324, 1988.

[29] Judy A Ungerer and Marian Sigman. Symbolic play and language comprehension in autistic children. *Journal of the American Academy of Child Psychiatry*, 20(2):318–337, 1981.

[30] Natalia M Kleinhans, L Clark Johnson, Todd Richards, Roderick Mahurin, Jessica Greenson, Geraldine Dawson, and Elizabeth Aylward. Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *American Journal of Psychiatry*, 166(4):467–475, 2009.

[31] Sally Ozonoff, Bruce F Pennington, and Sally J Rogers. Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of child Psychology and Psychiatry*, 32(7):1081–1105, 1991.

[32] Francesca GE Happé and Rhonda DL Booth. The power of the positive: Revisiting weak coherence in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, 61(1):50–63, 2008.

[33] Amitta Shah and Uta Frith. An islet of ability in autistic children: A research note. *Journal of child Psychology and Psychiatry*, 24(4):613–620, 1983.

[34] Kate Plaisted, John Swettenham, and Liz Rees. Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(5):733–742, 1999.

[35] Christopher Jarrold and James Russell. Counting abilities in autism: Possible implications for central coherence theory. *Journal of autism and developmental disorders*, 27(1):25–37, 1997.

[36] Therese Jolliffe and Simon Baron-Cohen. A test of central coherence theory: Can adults with high-functioning autism or asperger syndrome integrate fragments of an object? *Cognitive Neuropsychiatry*, 6(3):193–216, 2001.

[37] David Marr and Tomaso Poggio. From understanding computation to understanding neural circuitry. *A.I. Memo. Massachusetts Institute of Technology*, 1976.

[38] Jeffrey D Rudie, JA Brown, D Beck-Pancer, LM Hernandez, EL Dennis, PM Thompson, SY Bookheimer, and M Dapretto. Altered functional and structural brain network organization in autism. *NeuroImage: clinical*, 2:79–94, 2013.

[39] Brittany G Travers, Nagesh Adluru, Chad Ennis, Do PM Tromp, Dan Destiche, Sam Doran, Erin D Bigler, Nicholas Lange, Janet E Lainhart, and Andrew L

Alexander. Diffusion tensor imaging in autism spectrum disorder: a review. *Autism Research*, 5(5):289–313, 2012.

[40] Yanni Liu, Vladimir L Cherkassky, Nancy J Minshew, and Marcel Adam Just. Autonomy of lower-level perception from global processing in autism: Evidence from brain activation and functional connectivity. *Neuropsychologia*, 49(7):2105–2111, 2011.

[41] Marjorie Solomon, Sally J Ozonoff, Stefan Ursu, Susan Ravizza, Neil Cummings, Stanford Ly, and Cameron S Carter. The neural substrates of cognitive control deficits in autism spectrum disorders. *Neuropsychologia*, 47(12):2515–2526, 2009.

[42] Hideya Koshino, Rajesh K Kana, Timothy A Keller, Vladimir L Cherkassky, Nancy J Minshew, and Marcel Adam Just. fmri investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. *Cerebral cortex*, 18(2):289–300, 2007.

[43] Rajesh K Kana, Timothy A Keller, Vladimir L Cherkassky, Nancy J Minshew, and Marcel Adam Just. Atypical frontal-posterior synchronization of theory of mind regions in autism during mental state attribution. *Social neuroscience*, 4(2):135–152, 2009.

[44] Tal Kenet, Elena V Orekhova, Hari Bharadwaj, Nandita R Shetty, Emily Israeli, Adrian KC Lee, Yigal Agam, Mikael Elam, Robert M Joseph, Matti S Hämäläinen, et al. Disconnectivity of the cortical ocular motor control network in autism spectrum disorders. *Neuroimage*, 61(4):1226–1234, 2012.

[45] Daniel A Abrams, Charles J Lynch, Katherine M Cheng, Jennifer Phillips, Kaustubh Supekar, Srikanth Ryali, Lucina Q Uddin, and Vinod Menon. Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, 110(29):12060–12065, 2013.

[46] Daniel P Kennedy and Eric Courchesne. Functional abnormalities of the default network during self-and other-reflection in autism. *Social cognitive and affective neuroscience*, 3(2):177–190, 2008.

[47] Stewart H Mostofsky, Stephanie K Powell, Daniel J Simmonds, Melissa C Goldberg, Brian Caffo, and James J Pekar. Decreased connectivity and cerebellar activity in autism during motor task performance. *Brain*, 132(9):2413–2425, 2009.

[48] Natalia M Kleinhans, Todd Richards, Lindsey Sterling, Keith C Stegbauer, Roderick Mahurin, L Clark Johnson, Jessica Greenson, Geraldine Dawson, and Elizabeth Aylward. Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain*, 131(4):1000–1012, 2008.

[49] Michele E Villalobos, Akiko Mizuno, Branelle C Dahl, Nobuko Kemmotsu, and Ralph-Axel Müller. Reduced functional connectivity between v1 and inferior frontal cortex associated with visuomotor performance in autism. *Neuroimage*, 25(3):916–925, 2005.

[50] Yigal Agam, Robert M Joseph, Jason JS Barton, and Dara S Manoach. Reduced cognitive control of response inhibition by the anterior cingulate cortex in autism spectrum disorders. *Neuroimage*, 52(1):336–347, 2010.

[51] Dafna Ben Bashat, Vered Kronfeld-Duenias, Ditza A Zachor, Perla M Ekstein, Talma Hendler, Ricardo Tarrasch, Ariela Even, Yonata Levy, and Liat Ben Sira. Accelerated maturation of white matter in young children with autism: a high b value dwi study. *Neuroimage*, 37(1):40–47, 2007.

[52] Thomas W Frazier and Antonio Y Hardan. A meta-analysis of the corpus callosum in autism. *Biological psychiatry*, 66(10):935–941, 2009.

[53] Christopher S Monk, Scott J Peltier, Jillian Lee Wiggins, Shih-Jen Weng, Melisa Carrasco, Susan Risi, and Catherine Lord. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage*, 47(2):764–772, 2009.

[54] Christine Ecker, John Suckling, Sean C Deoni, Michael V Lombardo, Ed T Bullmore, Simon Baron-Cohen, Marco Catani, Peter Jezzard, Anna Barnes, Anthony J Bailey, et al. Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: a multicenter magnetic resonance imaging study. *Archives of general psychiatry*, 69(2):195–209, 2012.

[55] Lucina Q Uddin, Kaustubh Supekar, and Vinod Menon. Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in human neuroscience*, 7, 2013.

[56] David E Welchew, Chris Ashwin, Karim Berkouk, Raymond Salvador, John Suckling, Simon Baron-Cohen, and Ed Bullmore. Functional disconnectivity of the medial temporal lobe in aspergers syndrome. *Biological psychiatry*, 57(9):991–998, 2005.

[57] Christopher Lee Keown, Patricia Shih, Aarti Nair, Nick Peterson, Mark Edward Mulvey, and Ralph-Axel Müller. Local functional overconnectivity in posterior brain regions is associated with symptom severity in autism spectrum disorders. *Cell reports*, 5(3):567–572, 2013.

[58] Lonnie L Sears, Cortney Vest, Somaia Mohamed, James Bailey, Bonnie J Ranson, and Joseph Piven. An mri study of the basal ganglia in autism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 23(4):613–624, 1999.

[59] Adriana Di Martino, Clare Kelly, Rebecca Grzadzinski, Xi-Nian Zuo, Maarten Mennes, Maria Angeles Mairena, Catherine Lord, F Xavier Castellanos, and Michael P Milham. Aberrant striatal functional connectivity in children with autism. *Biological psychiatry*, 69(9):847–856, 2011.

[60] Charles J Lynch, Lucina Q Uddin, Kaustubh Supekar, Amirah Khouzam, Jennifer Phillips, and Vinod Menon. Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. *Biological psychiatry*, 74(3):212–219, 2013.

[61] Jose O Maximo, Elyse J Cadena, and Rajesh K Kana. The implications of brain connectivity in the neuropsychology of autism. *Neuropsychology review*, 24(1):16–31, 2014.

[62] Matthew K Belmonte, Greg Allen, Andrea Beckel-Mitchener, Lisa M Boulanger, Ruth A Carper, and Sara J Webb. Autism and abnormal development of brain connectivity. *Journal of Neuroscience*, 24(42):9228–9231, 2004.

[63] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411, 2013.

[64] Peter R Huttenlocher. *Neural plasticity*. Oxford University Press, 2002.

[65] Joan Stiles. *The fundamentals of brain development: Integrating nature and nurture*. Harvard University Press, 2008.

[66] Janet E Lainhart, Erin D Bigler, Maureen Bocian, Hilary Coon, Elena Dinh, Geraldine Dawson, Curtis K Deutsch, Michelle Dunn, Annette Estes, Helen Tager-Flusberg, et al. Head circumference and height in autism: a study by the collaborative program of excellence in autism. *American Journal of Medical Genetics Part A*, 140(21):2257–2274, 2006.

[67] David G Amaral, Cynthia Mills Schumann, and Christine Wu Nordahl. Neuroanatomy of autism. *Trends in neurosciences*, 31(3):137–145, 2008.

[68] Eric Courchesne, Ruth Carper, and Natacha Akshoomoff. Evidence of brain overgrowth in the first year of life in autism. *Jama*, 290(3):337–344, 2003.

[69] Yulia A Dementieva, Danica D Vance, Shannon L Donnelly, Leigh A Elston, Chantelle M Wolpert, Sarah A Ravan, G Robert DeLong, Ruth K Abramson, Harry H Wright, and Michael L Cuccaro. Accelerated head growth in early development of individuals with autism. *Pediatric neurology*, 32(2):102–108, 2005.

[70] Michael SC Thomas, Victoria CP Knowland, and Annette Karmiloff-Smith. Mechanisms of developmental regression in autism and the broader phenotype: a neural network modeling approach. *Psychological review*, 118(4):637, 2011.

[71] Rebecca P Lawson, Geraint Rees, and Karl J Friston. An aberrant precision account of autism. *Frontiers in human neuroscience*, 8, 2014.

[72] Elizabeth Pellicano and David Burr. When the world becomes too real: a bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10):504–510, 2012.

[73] Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1521):1211–1221, 2009.

[74] Andy Clark. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press, 2015.

[75] Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.

[76] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1):70–87, 2006.

[77] Karl Friston. The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, 13(7):293–301, 2009.

[78] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 1999.

[79] Pawan Sinha, Margaret M Kjelgaard, Tapan K Gandhi, Kleovoulos Tsourides, Annie L Cardinaux, Dimitrios Pantazis, Sidney P Diamond, and Richard M Held. Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, 111(42):15220–15225, 2014.

[80] Sander Van de Cruys, Kris Evers, Ruth Van der Hallen, Lien Van Eylen, Bart Boets, Lee de Wit, and Johan Wagemans. Precise minds in uncertain worlds: predictive coding in autism. *Psychological review*, 121(4):649, 2014.

[81] Shyam Sundar Rajagopalan and Roland Goecke. Detecting self-stimulatory behaviours for autism diagnosis. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1470–1474. IEEE, 2014.

[82] Chris M Bishop. Training with noise is equivalent to tikhonov regularization. *Neural computation*, 7(1):108–116, 1995.

[83] Ira L Cohen. An artificial neural network analogue of learning in autism. *Biological psychiatry*, 36(1):5–20, 1994.

[84] Kate C Plaisted. Reduced generalization in autism: An alternative to weak central coherence. 2015.

[85] Brian Egaas, Eric Courchesne, and Osamu Saitoh. Reduced size of corpus callosum in autism. *Archives of neurology*, 52(8):794–801, 1995.

[86] Andrew L Alexander, Jee Eun Lee, Mariana Lazar, Rebecca Boudos, Molly B DuBray, Terrence R Oakes, Judith N Miller, Jeffrey Lu, Eun-Kee Jeong, William M McMahon, et al. Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage*, 34(1):61–73, 2007.

[87] Michelle Luciano, Alan J Gow, Sarah E Harris, Caroline Hayward, Mike Allerhand, John M Starr, Peter M Visscher, and Ian J Deary. Cognitive ability at age 11 and 70 years, information processing speed, and apoe variation: the lothian birth cohort 1936 study. *Psychology and aging*, 24(1):129, 2009.

[88] Alan Feingold. Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, 62(1):61–84, 1992.

[89] Madhura Ingalhalikar, Alex Smith, Drew Parker, Theodore D Satterthwaite, Mark A Elliott, Kosha Ruparel, Hakon Hakonarson, Raquel E Gur, Ruben C Gur, and Ragini Verma. Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, 111(2):823–828, 2014.

[90] Richard Shillcock and Florian Bolenz. Male variability in general intelligenec. *submitted*, 2016.

[91] Laura S Allen, Mark F Richey, Yee M Chai, and Roger A Gorski. Sex differences in the corpus callosum of the living human being. *Journal of Neuroscience*, 11(4):933–942, 1991.

[92] Katherine M Bishop and Douglas Wahlsten. Sex differences in the human corpus callosum: myth or reality? *Neuroscience & Biobehavioral Reviews*, 21(5):581–601, 1997.

[93] Florian Bolenz. Modelling hemispheric interaction: Can connectivity explain differences in variability in general intelligence? *Unpublished Msc Dissertation*, 2016.

[94] Darold A Treffert. The savant syndrome: an extraordinary condition. a synopsis: past, present, future. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1522):1351–1357, 2009.

[95] Darold A Treffert. Savant syndrome: Realities, myths and misconceptions. *Journal of Autism and Developmental Disorders*, 44(3):564–571, 2014.

[96] Arthur Robert Jensen. *The g factor: The science of mental ability*. JSTOR, 1998.

[97] Christine N Vidal, Rob Nicolson, Timothy J DeVito, Kiralee M Hayashi, Jennifer A Geaga, Dick J Drost, Peter C Williamson, Nagalingam Rajakumar, Yihong Sui, Rebecca A Dutton, et al. Mapping corpus callosum deficits in autism: an index of aberrant cortical connectivity. *Biological psychiatry*, 60(3):218–225, 2006.

[98] Rich Stoner, Maggie L Chow, Maureen P Boyle, Susan M Sunkin, Peter R Mouton, Subhojit Roy, Anthony Wynshaw-Boris, Sophia A Colamarino, Ed S Lein, and Eric Courchesne. Patches of disorganization in the neocortex of children with autism. *New England Journal of Medicine*, 370(13):1209–1219, 2014.

[99] Giovanni Pezzulo, Francesco Rigoli, and Fabian Chersi. The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Frontiers in psychology*, 4, 2013.

[100] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[101] Therese Jolliffe and Simon Baron-Cohen. Are people with autism and asperger syndrome faster than normal on the embedded figures test? *Journal of Child Psychology and Psychiatry*, 38(5):527–534, 1997.

[102] Richard Zhang, Phillip Isola, and Alexei A Efros. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. *arXiv preprint arXiv:1611.09842*, 2016.

[103] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[104] Yoshua Bengio, Dong-Hyun Lee, Jorg Bornschein, Thomas Mesnard, and Zhouhan Lin. Towards biologically plausible deep learning. *arXiv preprint arXiv:1502.04156*, 2015.

[105] Benjamin Scellier and Yoshua Bengio. Towards a biologically plausible backprop. *arXiv preprint arXiv:1602.05179*, 914, 2016.