The Autism Palette:
Combinations of Impairments Explain the Heterogeneity in ASD

Ábel Fóthi[1,2], András Lőrincz[2*], Latha Soorya[3]

[1]Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
[2]Department of Artificial Intelligence, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary
[3]Rush University Medical Center, Chicago, USA

* Correspondence: Corresponding Author lorincz@inf.elte.hu

Keywords: neuropsychiatric disorder, cognition, social behavior, dimensionality, reward, heterogeneity, autoencoder, components, genetic causes

Number of words (without figure captions): 11568
Number of figures: 6
Number of references: 209

Abstract

Autism spectrum disorder (ASD) is a heterogeneous neuropsychiatric problem with a few core symptoms: weaknesses in social behavior, verbal impairments, repetitive behavior and restricted interests. Beyond the core symptoms, autism has strong association with other disorders such as intellectual disability, epilepsy, schizophrenia among many others. This paper outlines a theory of ASD with capacity to connect heterogeneous 'core' symptoms, medical and psychiatric comorbidities as well as other etiological theories of autism in a unifying cognitive framework rooted in neuroscience and genetics.

Cognition is embedded into an ever-developing structure modified by experiences, including the outcomes of environment influencing behaviors. We introduce the hypothesis that autism is caused by deficits in component-based cognition and the internal learning reinforcing machinery. Specifically, we outline our *Cartesian Factor* forming autoencoder like model that supports cognition by breaking combinatorial explosion and discuss the cognitive and neural processes behind our model.

The high dimensionality of sensory information poses serious problems, since the brain can handle only 7±2 relevant variables at a time making processes, such as the extraction and encoding of the relevant variables and their efficient manipulation critical. These processes are influenced by previous experiences and the internal reward system. In addition, large delays of distributed information processing should be counteracted by learned predictive models to synchronize sensory, proprioceptive, and cognitive signals and have timely and accurate model-based actions.

Impairments in any of these aspects may disrupt learning and execution. Combinations of small impairments may allow the solving of low complexity tasks but may become visible if learned variables and the related metric are improper and imprecise, respectively, especially if their number is large. We claim that social interactions are amongst the most challenging cognitive tasks in terms of the number of variables involved. In turn, they are highly susceptible to combinations of small impairments. We consider impairments as the basic colors of autism, whereas the combinations of diverse impairments make the palette of autism. In turn, social processes can be spoiled in many ways and can lead to diverse comorbidities.

1   Introduction

Although intellectual disability (IQ<70) is a common comorbid condition in autism spectrum disorder (ASD) and the mean IQ of autistic individuals is below 100, many individuals have normal or above average IQ (for specific data on prevalence, see, Baio et al, 2014). Moreover, only a proportion of people with intellectual disability (ID) are diagnosed with ASD. This variability in general cognitive ability is central to our unifying neurocognitive theory of ASD. Specifically, we propose that ASD results from impairments that corrupt learning and memory formation. In this theory, social cognitive processes are viewed as a particularly vulnerable cognitive mechanism due to the high complexity of social interactions combined with impairments in understanding intentions and emotion.

We will start with sensory information processing. Consider the millions and millions of sensory neurons. They give rise to an unimaginably huge space from the computational point of view. Since the number of sensors enters the exponent, the input is at odds with our cognitive capacity that can only deal with 7±2 items at a time. If the size of the space cannot be reduced, as we propose happens in various ways in ASD, then searching for solutions to problem solving and executive functions become troublesome in practice.

Concerning the reduction of information flow, the sheer number of variables required for success in a goal-oriented task deserves special attention. In IQ tests, mental manipulation of objects in matrix or verbal categorization tasks requires mostly the consideration of variables such as shape, orientation, weight, rigidity, color that typically directly influence visual, tactile and proprioceptive sensors, and only a few at a time. They are also simple in the sense that objects barely interact or change in time which considerably simplifies the complexity of the cognitive tasks (Fig. 1a).

By contrast, problem solving tasks related to 'social manipulations' are dynamic, complex and abstract. Human behavior changes quickly based on our own actions, or intentions of others towards ourselves or towards other objects. These 'parameters' are well hidden from direct observations. In addition, social interactions become more efficient when signals are provided, including utterances and gestures, tactile information that may also be the tools of deception (Fig. 1b). Changes in these information sources typically need highly adaptive and very quick responses. The fast synchronization of actions in the distributed and relatively slow neural system is not a trivial requirement and may be corrupted in many ways.

The paper is organized as follows. In the next section, we review relevant information theoretical concepts and introduce our model built on component forming autoencoding principles. We also show that genetic studies fortify our model. We consider local and global neuronal features of autistic individuals that can counteract component formation. Armed by these concepts and the experimental findings, we describe the autism palette, i.e., the different causes, or 'colors' that can lead to ASD when combined. We relate our model to other cognitive theories, comorbidities, neurobiological theories and review genetic underpinnings.
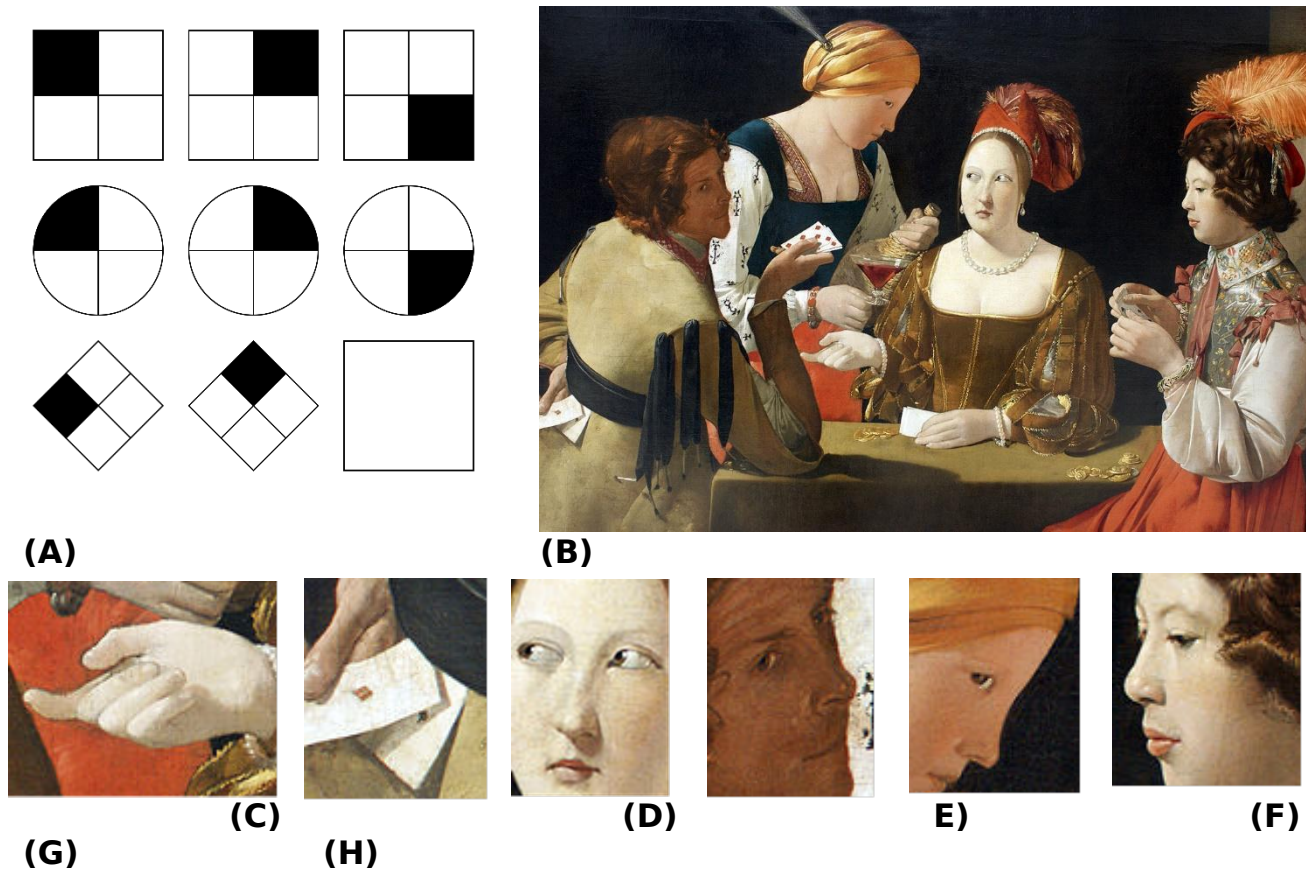
Figure 1. The difference between IQ tests and social interactions

(A) The variables: Shape of the object (square and circle) is irrelevant. Position of the black subfigure (clockwise rotation) is relevant. The subject is supposed to select the relevant variable(s) and to find the rule(s). The puzzle becomes harder very quickly, as the number of variables increases, and the rule is a combination of the variables. Probabilities, hypotheses are not involved. Reproduced from https://en.wikipedia.org/wiki/Intelligence_quotient under CC BY-SA 3.0 license

(B) The variables: the scenario for five people, their individual roles in the scenario, their relative positions in three dimensions to be estimated from a two-dimensional projection, occlusions of diverse objects and body parts for each person, the role of gaze directions (looking away, looking into the cards, looking to somebody), attention (seemingly no card related attention, attention to the own cards, attention to a person), objects in the hands (nothing, card or cards, glass of wine, bottle of wine), role of the hand (context restricted role of the hands within the context of the card game (showing the cards, hiding cards, covered pointing), the role of different cards in the game (ace versus other cards), role of people (players, made), the history of the people (sitting, walking) and the related potential observations (unobservable and observable objects for different partners), hypotheses about past observation (has seen something), uncertain observations on facial expressions that context can modify (compare **(B)** and **(E)-(H)** and the presumed goals (winning, making others to believe something, call somebody's attention to something, asking somebody in case of uncertainty) as well as the irrelevant variables related cloths, hair styles, jewelry, light intensities, amount of money on the table, including the varieties of forms, colors, shapes, and so on. Reproduced from https://www.wikiart.org/en/georges-de-la-tour/the-cheat-with-the-ace-of-diamonds under Public Domain license.

2   The Component Based Cognitive Model

Concepts developed for artificial neural networks serve us for modeling cognition. *Autoencoder* is the first concept. Autoencoders can emerge compressed predictive models of the world serving the *making sense process* of internal and external sensory observations.

2.1   Autoencoder

The autoencoder (Fig. 2a) is a computational architecture made of two types of mappings: a bottom-up (encoding) and a top-down (decoding) parts. Encoding maps the input to the so-called internal representation, while decoding uses that representation and produces an estimation of the input. Encoding, decoding and in turn, the internal representation itself is learned. The guiding principle is compression; it is easier to deal with a small number of representational entities than with the mass of sensory units. Internal representations can be predictive. In other words, interactions between the elements of the internal representation called the temporal propagation of the representation mimics sensory processes in such a way that decoding produces a faithful and timely match to the sensory information in spite of the processing delays (Fig. 2b).

Compressions come in different forms. One type is the so-called dense code. This case is like the zip code: components are interlinked, and no individual element of the code may make sense (Fig. 2a left hand side). Another version is called sparse code. This case differs: individual elements of the code may represent individual higher order correlations in the input, such as the edges of images (Olshausen and Field, 1996). Faces make another example. Sparse encoding produces a few of the indices, the code, of the representation, whereas decoding of one of the elements of the code may produce the nose and another element of the code may produce the mouth, and so on during the decoding procedure, see, e.g., Fig. 7 in (Makhzani and Frey, 2015). Due to the large varieties of eyes, noses, and mouths only a few of the indices represent any input and, in turn, the representation is 'sparse' (Fig. 2a right hand side). Furthermore, representations related to objects, such as houses, cars, cups, even paper clips (Bricolo, et al. 1997), or geographical locations (Moser, et al. 2008) may take negligible contribution if any in the decoding of a face. In this case, encoding selects a few units of the decoding network, producing a very compact representation of the input. Compression, i.e., the number of the active units that represent the input can be smaller in the sparse case than in the dense one due to the much larger number of units in the sparse case.

The smaller the number of the representational units that can provide a faithful estimation for an input, the better (i.e., the more probable) the internal model of the input space is. This is the so-called Occam's Razor Principle of information theory (Cover and Thomas, 2012) saying that "The simplest explanation is best". Sometimes, the content that a single element of the representation (i.e., a single decoding unit, or a component of the representation vector) holds / can produce / may emit is called a 'word' (i.e., a memory item), whereas the full set of the contents of the decoding units is called 'dictionary' (the set of memory items). Having this in mind, we say that sparse internal representation is the substrate that helps to 'make sense' of inputs in space and time by approximating (i.e., reconstructing or generating) the input by means of the words of the dictionary (Lőrincz et al., 2002). When the contents belonging to indices represent episodes, 'making sense' can concern past, present and future, i.e., a spatio-temporal model-based that predicts. Prediction of future inputs can be severely spoiled if some of the elements of the sparse dictionary are mixed or missing.
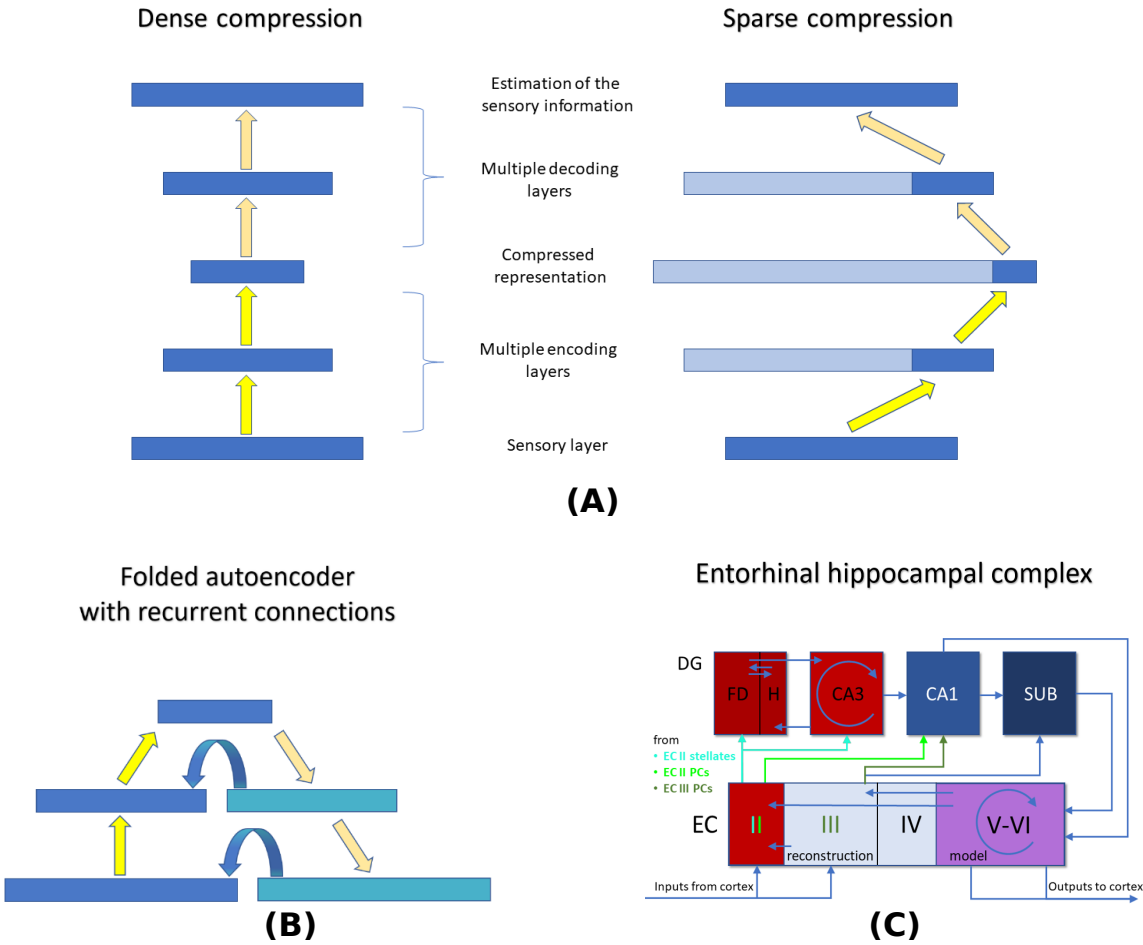
Dense compression　　　　　　　　　　Sparse compression

Estimation of the
sensory information

Multiple decoding
layers

Compressed
representation

Multiple encoding
layers

Sensory layer

(A)

Folded autoencoder
with recurrent connections

(B)

Entorhinal hippocampal complex

DG

FD　H　CA3　CA1　SUB

from
• EC II stellates
• EC II PCs
• EC III PCs

EC　II　III　IV　V-VI

reconstruction　model

Inputs from cortex　　　　　　　　　Outputs to cortex

(C)

6

Figure 2. Different compressing autoencoders and mapping to the neural substrate
Autoencoders compress inputs to representations and estimate inputs from representations. Sparse compression can be more efficient, i.e., fewer units may have to 'fire' at a time since 'firing units' may change from input to input. (**A** left-hand side): dense compression, (**A** right-hand side): sparse compression, (**B**): same as (**A**), but folded to form loops. Recurrent connections are added for the corresponding layers. The sparse version is similar. (**C**): sketch of the loop formed by entorhinal cortex (EC) and the EHC loop, a putative autoencoder, a crucial structure for forming and consolidating episodic memory. Main internal connections are shown. Superficial (deep) layers of the EC receive inputs (provide outputs) from (to) neocortical areas. Main stops of the information flow in the loop are the deep (V–VI) and superficial (II and III) layers of the EC, the dentate gyrus (DG), the CA3 and CA1 subfields and the subiculum. The loop in the DG connects the granule cells of the fascia dentata (FD) and the mossy cells of the hilus (H). The EC provides input to the CA3 subfield (perforant path and mossy fibers) and sends information to the CA1 subfield. Input also arises from EC III. Recurrent connections are present in EC V-VI and in CA3. PC: pyramidal cell. Recurrent connections make spatio-temporal model from the spatial one, propagate the representation forward in time for upstream estimation of the downstream signal and to compare them to produce errors for correction and for learning. See also Lőrincz & Buzsáki (2000), Chrobak, Lőrincz, & Buzsáki (2000), and Lőrincz (2018).

For complex and relevant actions, such as the treadmill stepping in infants (Yang et al., 2005), or the production of basic facial expressions of emotions (Hess and Thibault, 2009), partial genetic basis seems to exist. Still, the selection of the proper sparse representation and its synchronization with ongoing internal computations related to *concurrent episodic components*, including the elements of decision making, the launched actions that may occur 200 ms later and the sensory signals of the body and the environment that can be delayed by an additional 200 ms, are demanding and may be prone to errors.

Below, we elaborate on the concept of components in order to introduce our Cartesian Factors.

2.2   Components

Components in machine learning are identifiable sometimes autonomous parts of a larger system that can be hidden, may belong to an agent, or to its environment, e.g., the agent itself is a component of the environment. Components can be connected, may interact, and their coupling may give rise to reversible and irreversible changes in components. Typical component can be decomposed into smaller ones. An example is the eye, being part of the face and the body and having the iris, the retina and so on as its constituents. In the autoencoder, the indices of the representation and the contents belonging to these indices correspond to components.

In their simplest form, 'words' of the 'dictionary' correspond to the column vectors of a matrix (i.e., the dictionary) made of these columns; the columns are the memory vectors and the (active) indices of the columns make the representation. For the case of an image, columns of the dictionary are vectorized (i.e., concatenated pixels of image rows into a large vector), and the linear combination of some of these image patches make an estimation of the full image. In the form of a linear equation:

$$I(i,j) = \sum_{n=1}^{N} c_n A_n(i,j)$$

where i and j denote the $(i,j)^{th}$ pixel with value $I(i,j)$, c_n is the weighting factor of a column of matrix A, and $A_n(i,j)$ represents the nth column of this matrix having a vectorized form of the patch with pixels indexed by $i=1,…,I$ and $j=1,…,J$, and the number of columns is N. The width and height of the image and the patches are w and h with $0 < i < h$ and $0 < j < w$, respectively. In case if no confusion may arise, the word component may denote a component in the world, and an index in the representation, or the memory (e.g., the column vector) belonging to the index.

Recognition by components (Biederman, 1987) and holistic recognition help to understand the concept of 'words' and 'dictionary' and their hierarchical nature. For example, faces can be recognized holistically and through their components, see, e.g., the issue, edited by Watson and Robbins (2014) devoted to this subject in Frontiers of Psychology.

Components can be formed in diverse ways and psychologists suggested the Gestalt principles for learning the important ones a century ago (Köhler, 1967). Children's drawing support the theory both in terms of the 'words' / 'dictionary' and in the multi-level nature of the words themselves. Figure 3a shows a few examples from the 'Draw-a-Child' test representing components of the body and components of body components, such as hand and fingers, eyes and irises, among others. Similar principles have been used in computer vision, see., e.g., (Desolneux et al., 2004).
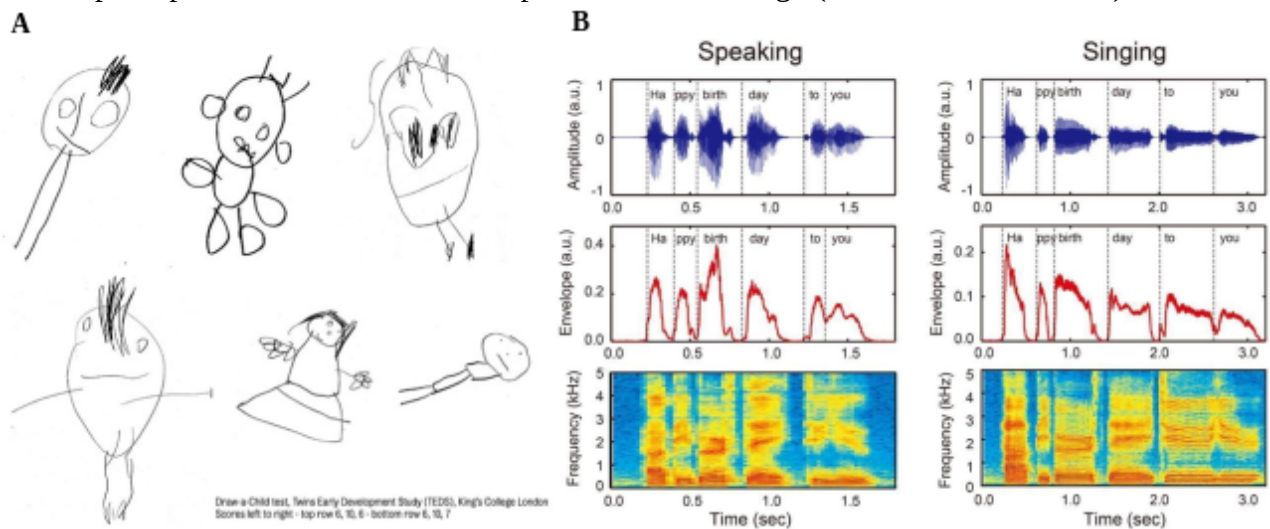

Figure 3. Visual and acoustic components
(**A**) Draw-a-Child test. Note the quality of the drawings and that many of the components of the body as well as components of the face are depicted in these drawings. Repreoduced with permission from King's College London. Their work was supported by the following grants: MRC grant G0901245 & NIH HD044454 & HD059215.
(**B**) Temporal components of the sentence "Happy birthday to you" in speech (left) and when singing (right). Acoustic components (phonemes) made visible by sliding window Fourier transformation in the temporal domain. From: Fujii & Wan (2014). Reproduced from [https://www.researchgate.net/publication/267741850_The_Role_of_Rhythm_in_Speech_and_La nguage_Rehabilitation_The_SEP_Hypothesis/figures?lo=1](https://www.researchgate.net/publication/267741850_The_Role_of_Rhythm_in_Speech_and_Language_Rehabilitation_The_SEP_Hypothesis/figures?lo=1) under CC BY 4.0.

However, components also exist in time as illustrated in Fig. 3b for the utterance 'Happy birthday to you'. Intense parts and more quiet regions follow each other, some of which are longer, others, like phoneme 'p' in *happy,* 'b' of *birthday,* and 't' that starts the word *to,* that is, the plosives are very short (on the order of tenths of milliseconds) compared to the vowels, and are not visible at the resolution of the figure.

Components come in a large variety, dictionaries differ from person to person, may change by time, and are influenced by external variables, such as light conditions, distance, occlusion and environmental noise, among others. In the vector-matrix case, contents of components of the representation are (non-linearly) summed up to approximate the input of the autoencoder.

Although this concept of components has similarities to our notion of Cartesian Factors (CFs) to be described below, it also differs from it. In particular, in many cases CFs don't sum up, but restrict space and time: they can be product-like (Friesen and Domingos, 2018). We believe that mammalian brain exploits both constructs. The joint set of product-like and summable components (i.e., CFs) seem to provide explanations for a number of ASD impairments and for social interactions, in particular.

2.3   Cartesian Factors

In this section we define our concept of components. There are many different formulations for 'factors' and 'components' in the literature, such as (1) non-negative components, (2) principal components, (3) independent components, among others, including their sparse multi-level (Diego and Hamprecht, 2018), and (4) non-linear extensions (Song, et al. 2018). In order to distinguish our concept from these, we introduce the concept of CFs that come in two different forms.

2.3.1 The Concept of Cartesian Factors

Type 1 CF, or CF1 for short is like traditional components. The main feature is that they are decomposable and are decomposed from larger systems. They are like Lego elements: they can be separated, and they can be put together.

Type 2 CF (CF2), differs. Three particular examples for CF2 are color, the three-dimensional space, and the one-dimensional time. Such CF2 features serve us to explain synesthesia, for example.

Type 2 CFs are special in the sense that roughly speaking, they don't exist by themselves. They can be modified only and can't be separated. They concern the type of the components under consideration. For example, any color is always bound to something. This is not the case for objects or episodes, since objects exist, episodes happen, and they occupy a local region in space and time, i.e., they exist *somewhere* and *only for a while* and are parts of concurrent processes. By contrast, CF2s (e.g., color) are not limited either by space or in time. A somewhat similar concept in philosophy is called qualia see, e.g., Jackson (1978) and the Stanford Encyclopedia of Philosophy. Typical examples for qualia are the blueness of the sky and the scent of a flower, (Frankish, 2016), and (Chrisley and Sloman, 2016), respectively.

We add the perception of the three-dimensional (3D) nature of the world, the feeling of acceleration, our emotions and pains of different kinds. Although the concept of qualia is a controversial concept (Kind, 2001), it may help to clarify the differences between CF1 and CF2s:

- We have access to CFs via our sensory-motor system
- CF1s can be added/coupled may interact and are decomposable components.
- CF2s don't exist by themselves, they are restrictive (product-like), they can be modified. CF2s are non-decomposable components.
- There are CFs that may correspond both to CF1s and CF2s; the two types of our Cartesian Factors are not exclusive. For example, the stripes on a horse-like creature can be interpreted as both CF1 (at the pigment level) and CF2s (as colors).
- CF2s may be decomposed, like color space can be decomposed by a prism, space into regions, time into intervals, and so on.

As a specific example that may take the reader closer to our concept of CF2, consider maze dimensions in different rat experiments: the constrained bottom region in a vertical rotating wheel is 'point-like', a radial arm maze is composed of a set of straight one-dimensional (1D) line segments, the Morris water maze is of two-dimensional (2D) nature, whereas bats live in a three-dimensional (3D)space (Heys et al., 2013). Such spatial descriptors belong to CF2s.

Frankish (2016) relates qualia to illusionism. Indeed, there are illusions that support the concept of both type of CFs. For example, component like form plays a role already in early visual processing: images with Gabor filters provide a good example (Ishai and Sagi, 1997). Having two Gabor filters of the same length set along a line, separated by their length, the sensitivity of observing a low contrast Gabor filter between them with the same orientation increases. One may say that lines are made of line segments and the missing segment (here, a segment component) is inferred by pattern completion mechanisms. Researchers reported a similar effect after imagining, i.e., without direct observation of the two separated Gabor filters indicating that an autoencoding mechanism is in effect in the visual stream.

The well-known Kanizsa illusion is somewhat similar. PacMan like figures 'define' an object that we see in a somewhat different color, but the color is the same as in its immediate neighborhood (Fig. 4a). This is also pattern completion that involves illusory contours.

Another illusion is qualia-like and is shown in Fig. 4b. Ends of line segments define a circle and the full disk defined by this circle gains the illusory color of the edges although the background is white. Possibly, color and shape are encoded by separate CF2 channels, namely, shape and color, whereas decoding joins them and joining can be imperfect.
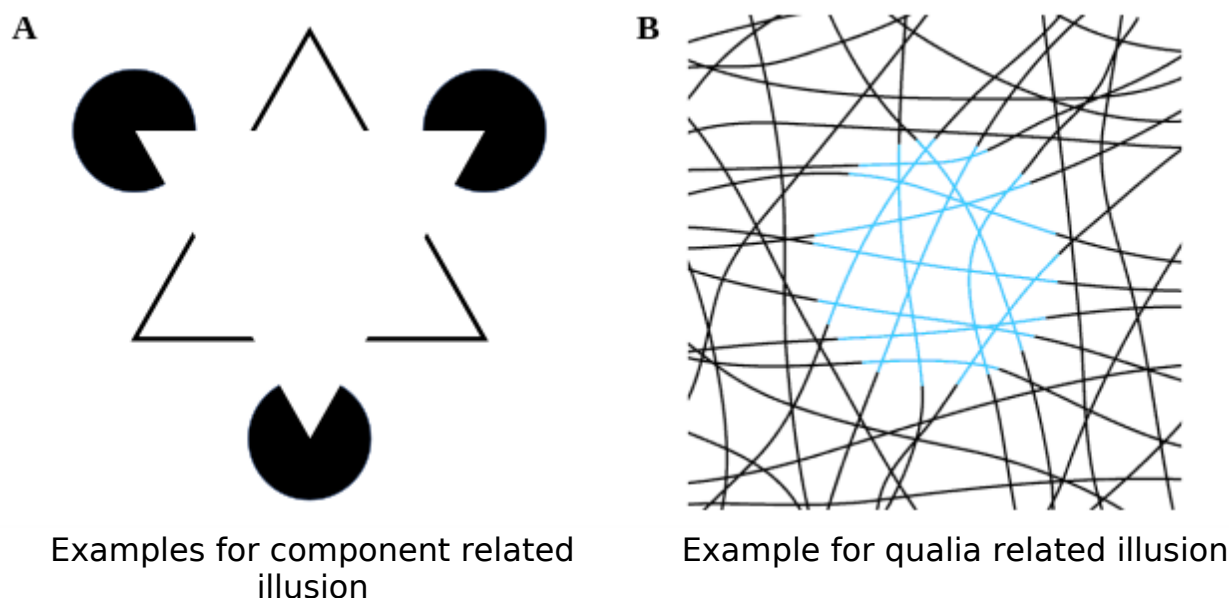
Examples for component related illusion

Example for qualia related illusion

Figure 4. Illusions
(**A**) Examples for component related illusion: Kanizsa triangle illusion. White color is identical everywhere. Pattern completion happens at the edges. Reproduced from https://en.wikipedia.org/wiki/Gaetano_Kanizsa#/media/File:Kanizsa_triangle.svg under CC BY 3.0 license
(**B**) Example for qualia related illusion: Neon color spreading illusion (lines are black and blue). There are only three colors – white, black and light blue – in the figure. Small rotation of the disk region that looks light blue removes the illusion, leaving only light blue lines against a white background. (see at https://michaelbach.de/ot/col-neon/index.html). A pixelwise examination of the image reveals the true colors. Reproduced from
https://commons.wikimedia.org/w/index.php?curid=29960445 under CC BY-SA 3.0 license.
Original source: http://www.blelb.com/english/blelbspots/spot05/images/neon07.gif

In spite of the large differences between mammalian species, some of the CFs, like 2D or 3D spaces and the belonging metric develop similarly in rats, bats, non-human primates and humans, see, e.g., (Heys et al. 2013). Although the algorithmic and representational details of CF1 and CF2 formation remain unclear, but many, if not all of them fall under the category of declarative memory. In turn, the entorhinal-hippocampal complex plays a critical role in CF formation. We shall return to this point later.

We add that CF2s may depend on culture and on scientific advances as demonstrated by (Zhong et al. 2018), for example.

2.3.2 Emotion as a Cartesian Factor

Emotions and the emotions of the partner(s) also form CFs, but unlike color, shape and dimensions, this CF space is well hidden. Emotion related observations can differ from individual to individual, and due to the well-hidden feature of these CFs, large deviations in the estimation of other person's emotions can be expected even for normal subjects, especially with different cultural backgrounds, not to mention individuals with ASD.

The visible signs of emotions are multi-faceted and include eye movements, body language, verbal expressions, and alike, and the range of any given emotion can vary from person to person. Furthermore, the duration of emotions can span over shorter and longer time periods, with some emotion cues occurring in a flash. Combinations of visible gestures may change from time to time and may be the subject to self-control.

The "emotional space" can be approximated in 2D, see the 'emotion wheel' of Plutchik (1980). Dynamic emotional features have been attempted to be captured by a metric based on combinations of muscle contractions and expansions, such as the facial action coding system (Ekman, 1982).

2.3.3 Complexity of Behavior and Social Learning

Figure 5 depicts the differences between factor searches in IQ tests and factors that influence social learning, including the time constraints.
We recall the arguments that
      a) Cartesian Factor formation is crucial for social interactions,
      b) the quality of social interactions can be influenced by many internal behavioral facets, including motivations, fear, efficiency of reinforcements, among others,
   c) the number of factors in social interactions can be very high (Fig. 1).
In addition,
      d) learning of social interactions relies on reinforcements

In turn, complexity is high due to the number of the involved factors and the additional burden of the temporal credit assignment problem in reinforcement-based learning (Sutton, 1985) that may need a large number of experiences, see later.

Components are useful for (a) pattern completion by adding missing CF1s in space and similarly, one can consider prediction as pattern completion in time, (b) generalization via eliminating many details, including some of the CF2s, and (c) selecting only a few of the many CFs for decision making. An example is the shape and texture descriptors of mushrooms to tell apart poisonous and edible ones. Decision time is not limited. In contrast with this mushroom example, social decision making is much more complex and decision time is severely constrained.

Reaction time: minutes
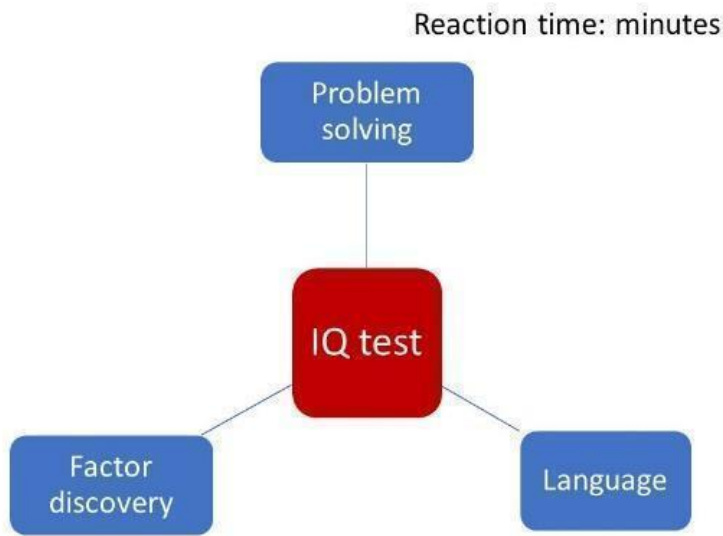
(A) Components needed for solving most IQ tests

Reaction time: seconds
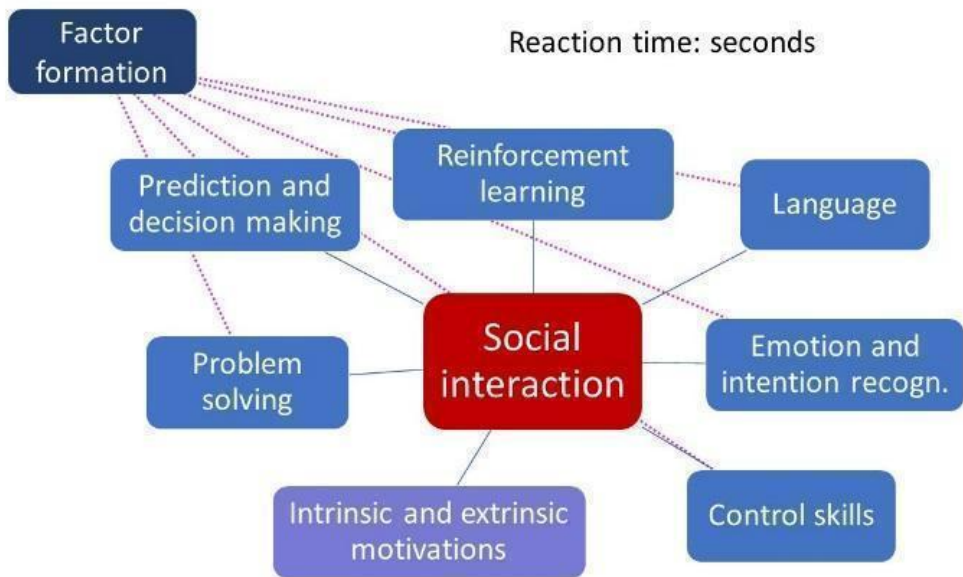
(B) Components needed for solving goal-oriented tasks in social interactions

Figure 5. Time constraints and components needed for solving (**A**) IQ tests and (**B**) goal-oriented tasks in social interactions. Solid lines: main components, dashed lines: dependencies from the point of view of factor formation.

In the brain, many details of CF formation have been uncovered at least for the representation of space (Moser, et al. 2008) and there are many theories on how CFs are formed. We shall consider these issues later.

2.3.4 Dimension Reduction for Behavior Optimization: Factored Reinforcement Learning

We argued that the optimization of behavior can be much harder than problems in IQ tests, since many variables in behavior optimization and social learning are well hidden. An additional issue is that rewards are delayed. In turn, reinforcement learning, this trial-and-error method should be invoked and the collection of any "learning sample" needs considerable time. Furthermore, reinforcement taxes the learning process, since the complexity of solving such problems is proportional to the number of variables *multiplied by three* (Kearns and Singh, 2002) and this number enters to the exponent.

If the few, but necessary factors are given for each step of decision making then one talks about factored reinforcement learning (Kearns and Koller, 1999) that exhibits favorable scaling properties even if there are many steps and the number of all the variables is large (Szita and Lőrincz, 2009). However, theoretical solution for the joined problem, i.e., factor learning *and* behavior optimization, are needed for the optimization of human-machine interactions but does not exist as of yet.

Missing information is another serious obstacle according to the theory of reinforcement learning (Krishnamurthy, 2016), since in such cases the actual state becomes uncertain. CF1s can approximate missing information pieces and thus to simplify the learning of behavior optimization.

2.4   Cartesian Factors and Autism are Deeply Rooted in Genetics

Dimension reduction is a well conserved mechanism, which is absolutely necessary for actions and intelligence. Dupre and Yuste (2017) showed that even hydras, a genus of cnidarians, are capable to reduce the inputs from the outside world to four components. The manipulation and combination of these components make them able to carry out basic behavior. Both the concept of neuronal circuits and the underlying molecules and genes are similar in all animals from hydras to humans. In the hydra neural network, the four components are manifested as four distinct clusters of co-activated neurons. These clusters are anatomically non-overlapping functional circuits and their neurons are interacting with each other by sodium, potassium, and calcium channels and receptors for glutamate, GABA, dopamine and other conserved peptides.

2.4.1 Polygenic Model of ASD

Recent results show that the development of ASD on the molecular level mainly determined by genetics and partially by epigenetics. Twin studies estimate ASD heritability to be approximately 80-90% with additional small, but significant environmental influence (Colvert et al., 2015; Sandin et al. 2017; Tick et al., 2016). Genetics also supports our view about the heterogeneous sources of ASD. From the point of view of genetics, autism is a complex disease. In contrast to Mendelian diseases, where a single causative variant in a single gene can be pinpointed, complex diseases are the outcome of a plethora of variants with variant effect sizes. Based on genetic linkage and genome-wide association studies (GWAS) the number of ASD candidate genes is around 1000 (Devlin and Scherer, 2012). These genes have various functions, divergent expressional patterns dependent both on tissue type, condition and developmental phase. Furthermore, genetic variants found in GWAS are mostly non-coding variants (Buniello et al., 2019) with small influences on

autism pathogenesis. Genetic variants have variable penetrance, most of them being low or unknown (Vorstman et al., 2017). In about 90-95% of the ASD population the causative variants are unknown or predicted to be common, inherited variants.

These findings support the polygenic risk model, where the development of the disease is the result of the combination of genes with small individual effect on ASD pathogenesis. In these cases, there are thousands of causative genetic variants and several hundreds of affected genes (Gandal et al., 2018).

In about ~5% of the cases are monogenic forms of ASD (de la Torre-Ubieta et al 2016), where a single genetic source can be pinpointed. However, this genetic variation is rarely inherited, mostly *de novos* and are on the chromosomal level, thus affecting several genes at the same time. A large portion of the monogenic cases are among the syndromic forms of ASD, where the number of candidate genes can range from a single gene to hundreds of genes (De Rubeis et al., 2014). Among these genes pleiotropic genes are frequently involved. Pleiotropy describes the ability of a gene to influence multiple traits or diseases (Solovieff et al. 2013).

Conclusively, the heterogeneity of ASD can be explained by polygenicity and the pleiotropic nature of ASD candidate genes (Smoller et al., 2013). In light of polygenicity and pleiotropy, frequent comorbidity of other psychiatric and neurological diseases with ASD makes sense.

2.4.2 Molecular Pathways and Neural Circuits as Vulnerabilities

The high number of putative genetic variants and the pleiotropic nature of affected genes shifted the interest of geneticists to molecular pathways, where the effect of multiple genes can be integrated, thus the number of causative molecular mechanisms can be reduced (Pinto et al., 2014). In ASD research, the most interesting pathways are: 1) those that regulate cell growth, 2) connect cells to each other and 3) make possible for them to communicate, form circuits and regulate the network of these circuits.

From the point of view of neuropsychiatry, dysfunction in an ASD related pathway may be considered a vulnerability, which in combination with other vulnerabilities can manifest in the development of ASD (Beauchaine and Constantino, 2017). As pleiotropic genes are active in multiple pathways there are also pathways that can influence multiple diseases, as in the case of neuropsychiatric diseases, where shared pathways are frequent (Gandal et al., 2018) and they can be seen as shared vulnerabilities of comorbid diseases. Further, the severity of ASD is presumed to be dependent on the type of genetic variants and the number of impaired vulnerabilities (van de Lagemaat and Grant, 2010). For example, genetic variants with cell type or developmental phase specific effects (e.g., non-coding single-nucleotid variants) may have a weaker effect than variants with more general influence (e.g., chromosomal anomalies, copy number variants, likely gene-disrupting mutations).

2.4.3 *mTOR* Pathway

One of the most studied pathways is the *mTOR* pathway, which is strongly affected in multiple traits/diseases: e.g. cancer (Efeyan and Sabatini, 2010), epilepsy (Lipton and Sahin, 2014), intellectual disability (ID) (Troca-Marín et al. 2012), ASD (Kelleher and Bear, 2008), and related phenotypes: e.g. macrocephaly (Lee et al., 2012). Mutations in factors of the mTOR pathway dysregulate neuronal translation. Both of the above-mentioned neurological diseases show high

overlap with ASD: ~ 30-40% of patients with ASD also have epilepsy (Amiet et al., 2008; Yasuhara, 2010) and they frequently have brain overgrowth (Lainhart et al., 1997). Comorbidity with epilepsy also increases the risk of ID and the prevalence of epilepsy and ID, both is higher in autistic females than males (Amiet et al., 2008). The overactivation of *mTOR* pathway can cause epilepsy through increasing spine density and excitation (Li et al., 2010), but on the other hand it can induce macrocephaly by increasing neuronal size (Crino, 2011). 30-40% of autistic children experience brain overgrowth in their early childhood (Courchesne et al., 2001; Lainhart et al., 1997) and in cases of high-risk children this ratio is even higher (Hazlett et al., 2017a). Some ASD candidate genes are implicated in brain size regulation in a dose-dependent manner (Horev et al., 2011). Knowing that *mTOR* pathway regulates several important cell functions such as cell proliferation, synaptogenesis (Crino, 2011), and variants in the regulators of *PI3K-AKT-mTOR* pathway were detected in half of the cases of co-occurring macrocephaly and ID (Yeung et al., 2017), impairment in *mTOR* pathway is the most probable mechanism behind ASD related macrocephaly. Furthermore, loss of *FMR1* (key gene in Fragile X syndrome) also enhances *mTOR* activity (Sharma et al., 2010). Three genes among the early discovered ASD genes (*PTEN* and *TSC1/TSC2*) are repressors of *mTOR* (Inoki, et al. 2005). In *TSC1/TSC2* knock-out (KO) inhibition decreases, while in *PTEN* KO both inhibition, excitation and spine density increase (Gao and Penzes, 2015). Both of these can lead to elevated E/I ratio and epilepsy (Bateup et al., 2013). Altogether, loss of function of either *PTEN* or *TSC1/TSC2* enhances *mTOR* activity, but in different ways and with different co-functions. Enhanced *mTOR* activity is a severe impairment with a general effect on brain development and a cause of several overlapping co-morbid outcomes.

2.4.4 Synaptic Impairments can Disrupt ASD Related Circuits

Besides direct and indirect regulators of neural growth, many ASD candidate genes are factors of molecular pathways with synapse specific functions (Bourgeron, 2009) such as receptors, ion channels, adhesion molecules, and other membrane and scaffold proteins. Excitatory and inhibitory synapses are characterized by different sets of molecules. Pre-synaptic vesicular glutamate transporters (VGLUTs) and receptors (NMDARs and AMPARs) are specific for excitatory synapses. Inhibitory synapses have GABA transporters (VGATs) and receptors (GABAARs and GABABRs) (Eccles, 2013). Their dysfunctions lead to excitatory and inhibitory imbalance, an emerging hypothesis of ASD pathomechanism (Gao and Penzes, 2015). In case of synaptic genes, a neuronal specific form of molecular vulnerabilities, the circuitry becomes involved. While biochemical pathways (e.g. the previously mentioned *mTOR* pathway) are mainly defined intracellularly, circuits are describing intercellular connections, which are fulfilled through synaptic communication.

Synaptic genes are good examples to show the specific impact of different genes with similar molecular function. *SHANK*s, a family of scaffold proteins, fulfill important adapter function by anchoring and clustering receptors, ion channels and other membrane proteins on the postsynaptic membrane (Sheng and Kim, 2000). They are among the most studied ASD genes and influencing ASD severity by distinct cognitive impairments. Deletion, truncation, or loss-of-function of *SHANK3* viewed as the most important cause of Phelan-McDermid Syndrome, which is highly comorbid with severe ID (Bonaglia et al., 2001; Wilson et al., 2003), while defects of *SHANK2* and *SHANK1* are enriched in ASD with mild ID or without ID, respectively (Leblond et al., 2014).

Neurexins and neuroligins are important adhesion molecules and strong candidate genes of ASD and other cognitive diseases (Südhof, 2008). They stabilize synaptic connections and shape them by tuning their properties, like their excitatory/inhibitory activity (Chih, et al. 2005). Their function can be very specialized due to their huge number of splicing variants (Ichtchenko et al., 1996).

Formation of neuronal circuits is heavily dependent on the precise regulation of alternative splicing of these adhesion molecules (Nguyen et al., 2016).

Here we have reviewed that the diverse genetic background behind ASD can induce heterogeneity through affecting several neuronal pathways: neuronal growth, synaptogenesis, synaptic plasticity and many others. ASD symptoms and comorbidities could depend on the number and the type of impairments in these pathways. Some autism related genetic variants can severely damage intelligence as we would expect from impairments of CFs, while others do not influence intelligence at all. We suggest that these variants affect reinforcement and we discuss them in the next section; we extend the theory supporting molecular findings with neuropeptides and neurotransmitters.

2.5   Neurotransmitters in Social Behavior and Reward
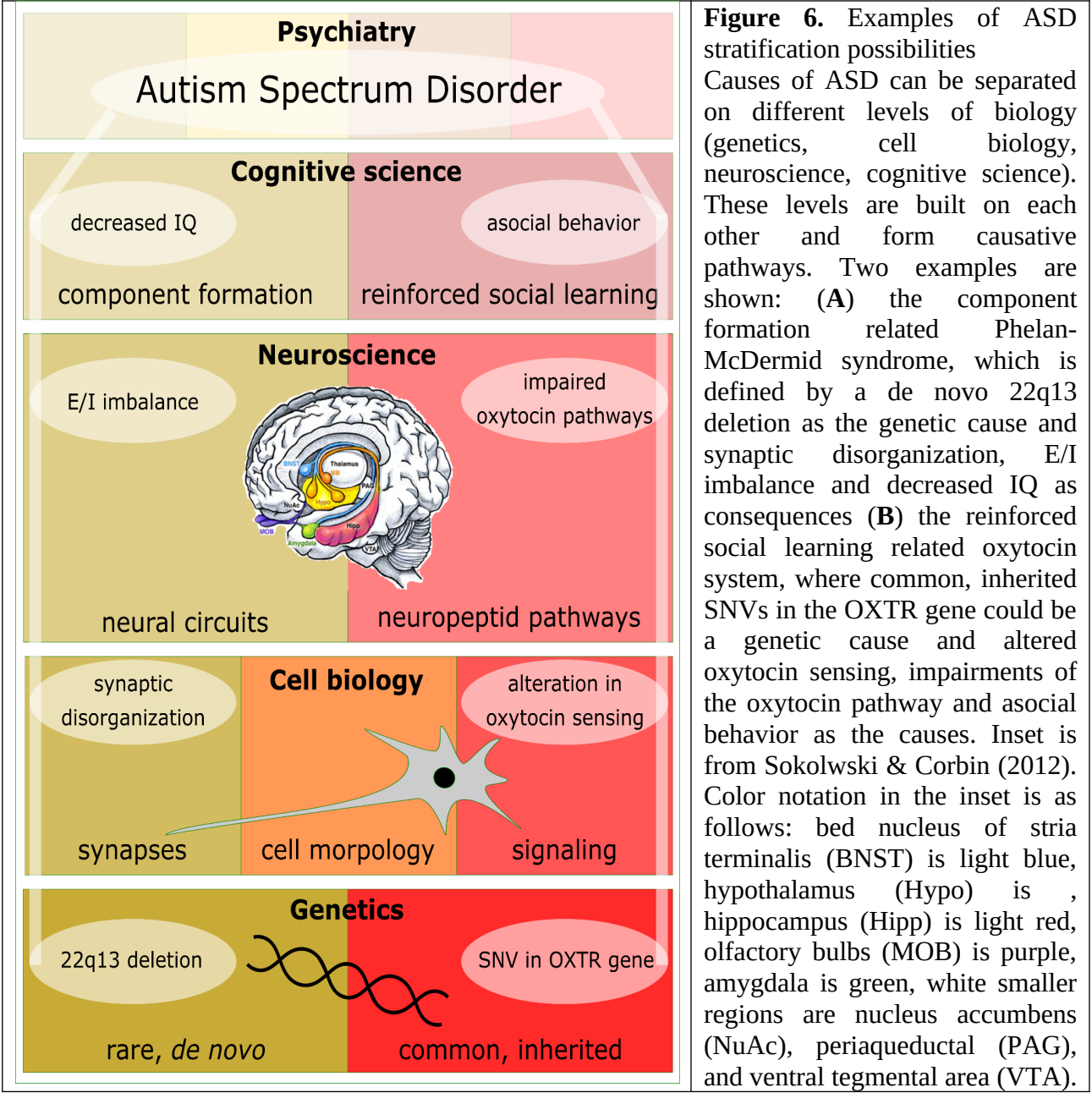
2.5.1 Oxytocin

Oxytocin (OXT) can affect a network of circuits and can coordinate complex brain functions to form specific behavioral phenotypes. Function of OXT is intensively studied in social behavior and autism. From recent works a specific network of neural circuits can be concluded. OXT realized from the hypothalamic paraventricular nucleus promotes prosocial behavior by increasing the excitability of the dopaminergic neurons in ventral tegmental area (VTA), which are projecting to the nucleus accumbens. This circuit is a positive, *rewarding loop for social interactions* (Hung et al., 2017). These projections are part of the mesocorticolimbic dopamine (DA) pathway, a network of circuits, which was shown to be impaired in ASD in multiple studies (Ernst et al., 1997; Supekar et al., 2018). In contrast to the positive effect of OXT on DA neurons of VTA, OXT indirectly inhibits DA neurons in substantia nigra, hence suppressing exploratory locomotion (Angioni et al., 2016; Xiao et al., 2017). Impairments in the substantia nigra were implicated as *cause of repetitive behavior* in ASD (Fuccillo, 2016; Kim et al. 2016). These results connect the above mentioned OXT and DA circuits to two of the core behavioral symptoms of ASD. Further, the importance of this circuits is supported by the fact that an abnormal social interaction can be triggered by VTA DA neuron specific deletion of OXT receptor (Hung et al., 2017) or *neuroligin-3* (Bariselli et al., 2018), an ASD-related synaptic adhesion molecule. Moreover, mimicking an ASD related loss-of-function mutation of *neuroligin-3* in rodents, repetitive behavior and aggression increase, which can be reversed by Risperidone, a D2 DA antagonist (Burrows et al., 2015). These receptors and adhesion proteins are coded by genes, affected by genetic variants, they are factors of molecular pathways and their deregulation can disrupt neuronal circuits that have importance in ASD related behavioral phenotypes. Altogether these examples illustrate certain bridges between genes and phenotypes, shown in Fig. 6.

2.5.2 Dopaminergic Pathway

It has been known for more than 20 years that the dopaminergic pathway plays a role in reinforcement learning (Ljungberg et al. 1992; Waelti et al., 2001). In turn, pathway impairments will spoil the optimization of behavior and, in particular, social learning.

Beyond impairments in social learning, this pathway influences the learning of motor control. For a theoretical description and the review of the neuronal findings, see, e.g., (David et al. 2008; Lőrincz, 2018) and the references therein. Experimental results also show the complexity of potential genetic influences: one gene plays a role in probabilistic reward learning and another one

in avoiding choices that give rise to negative outcomes stochastically, whereas a third one is involved in the ability of rapid adaption from trial to trial (Frank et al. 2007). The contribution of this pathway in autism has been demonstrated in the literature, see, e.g., (Kruppa et al., 2019; Larson et al. 2011), and the cited references.

Figure 6. Examples of ASD stratification possibilities

Causes of ASD can be separated on different levels of biology (genetics, cell biology, neuroscience, cognitive science). These levels are built on each other and form causative pathways. Two examples are shown: (**A**) the component formation related Phelan-McDermid syndrome, which is defined by a de novo 22q13 deletion as the genetic cause and synaptic disorganization, E/I imbalance and decreased IQ as consequences (**B**) the reinforced social learning related oxytocin system, where common, inherited SNVs in the OXTR gene could be a genetic cause and altered oxytocin sensing, impairments of the oxytocin pathway and asocial behavior as the causes. Inset is from Sokolwski & Corbin (2012). Color notation in the inset is as follows: bed nucleus of stria terminalis (BNST) is light blue, hypothalamus (Hypo) is , hippocampus (Hipp) is light red, olfactory bulbs (MOB) is purple, amygdala is green, white smaller regions are nucleus accumbens (NuAc), periaqueductal (PAG), and ventral tegmental area (VTA).

2.6 Some Local and Global Neuronal Features and Their Relevance in Autism

In this section, we touch upon some features of the neuronal system that contribute to factor formation and are relevant in our approach to understand autism. We bridge this section to the

previous ones by a short outline of our predictive autoencoding view of the brain serving goal-oriented behavior.

2.6.1 Factor Formation

The two types of CFs can develop in different ways as shown by computational models. Component-like CFs (CF1s) may be formed by Gestalt-like mechanisms by means of statistical information using similarity measures both in space and time. A low-level spatial autoencoder model (Olshausen and Field, 1996) developed local filters similar to those found in the primary visual cortex. Spatio-temporal hierarchical autoencoder models of similar kind have also been formulated, see, e.g., Milacski et al. (2019), and the cited references therein. Both methods develop sparse representations in the form of a low-level dictionary for decoding. Novel works emphasize the grammar like nature of visual information (Baraniuk et al. 2010; Friesen and Domingos, 2018).

Learning of certain CF2s seem to occur in the loop of the enthorinal-hippocampal complex (EHC loop). We sketch the example of space. Space is related to control, including navigation, and the metric appears to be encoded in the medial entorhinal cortex (see the book of Andersen et al., (2006) and the cited references therein). The space related code (kind of a soft discretization) is in the hippocampus that has place cells firing in local regions independently from the head direction, light conditions, and some other attributes. Neurons related to the metric of the space fire along triangular grids of two-dimensions in the entorhinal cortex in rodents. Both grid cell and place cell firings correspond to sparse population coding and the representation provided by place cells is sparser than that of grid cells (Quiroga and Kreiman, 2010).

Triangular organizations concerning the representation of the metric of a specific 2-dimensional cognitive task (Constantinescu et al. 2016) have been found in the neocortex. In this particular case, the 2D space is formed by the lengths of the neck and the legs of birds.

Neurobiology indicates that CF2s related to space are formed by semi-supervision, i.e., an internal signal supervises learning: head direction cells may develop from vestibular inputs and self-motion cues among other information sources and such head direction cells play a fundamental role in the learning process that forms place and grid cells (Taube, 2017; Winter and Taube, 2014; Cullen and Taube, 2017). Lesioning the head direction cell system spoils both place cell and grid cell formations. A sparse autoencoding model successfully developed place cells and directed grid cells from visual input and head direction supervisory signals (Lőrincz and Sárkány, 2017). It has been shown that direction independent grid cells can be developed by means of reinforcement learning (Banino et al., 2018).

2.6.2 Brain Waves

As mentioned before, the EHC loop is a key component of learning of episodic and semantic memories. Brain waves play an important role in these processes. Buzsáki (1989) suggested that during exploration, i.e., during the theta waves, neocortical information enters the hippocampus and induces weak and transient heterosynaptic potentiation in the CA3 subfield. Later, these weakly potentiated neurons and neuronal chains cause sharp waves (SPW) during consummatory behaviors that give rise to long-term synaptic modifications in the hippocampus. Long-term encoding of these memories involves complex patterns of brain waves.

According to the experimental results, SPW and slow-wave sleep (SWS) play roles in the consolidation and synchronization processes. SWS seems to be more involved than REM sleep in the consolidation process, but both types of sleep contribute (Born and Wilhelm, 2012; Diekelmann and Born, 2010). SWS or non-REM sleep is characterized by joint spindle and ripple events. During these events, the depolarizing up phases of slow oscillations reactivate the hippocampal time series and SPW ripples. Different brain wave frequencies can couple, and slow waves seem to restrict the time interval for coupling when faster oscillation occurs within slower ones, a phenomenon called phase biasing. In turn, slow oscillations can coordinate local processes globally, making SPW ripples a highly synchronous wave pattern in the brain. Boyce et al. (2016) showed causal evidence that contextual memory consolidation depends also on theta rhythms during REM sleep. Errors of the brain-wave machinery may have strong effects on the forming and the synchronization of the component-based memory system.

We turn to the autoencoder model inspired by the features of encoding space and the related metric in the EHC loop.

2.6.3 Sparse Autoencoders in the Brain

It is customary to say that an autoencoder is 'dreaming' when decoding is driven by the freely running predictive networks (Fig. 2b) without any input. In turn, autoencoders can support learning in two ways: (i) the error of the estimated input can improve the representation (see, e.g., Goodfellow et al., 2015) and the cited references therein. This learning is Hebbian in the folded architecture (Lőrincz and Buzsáki, 2000). (ii) Certain dreaming phases may consolidate memory traces of behaviorally relevant parts by rehearsing them (Crick and Mitchinson, 1983). A large amount of research has been devoted to memory consolidation in the brain (see, e.g., Watson and Buzsáki (2015)) catalyzed by the two-stage brain waves mechanism (Buzsáki, 1989) described above.

It has been suggested that the EHC loop forms a predictive autoencoder (Lőrincz and Buzsáki, 2000; Chrobak et al., 2000) that (i) can learn from errors of the estimations via the two-stage operation, (ii) can learn and compensate for delays, and (iii) can develop independent components (Fig. 2c). This line of research has been progressing: metric learning, learning of episodic representations (Szirtes and Lőrincz, 2009), and semi-supervised learning of *sparse* Cartesian Factors (Sárkány and Lőrincz, 2017) have been included. The architecture is depicted and detailed in Fig. 2c. Findings on delay compensation (Henze et al., 2002) reinforced the model, whereas others on stellate and principal cell projections from the entorhinal cortex to the hippocampus reviewed in (Valero and Menendez de la Prida, 2018) may serve to sharpen it (Lőrincz, 2019).

An extreme version of sparse representation is the concept of grandmother cells, that is a cell that responds only to the image of one's grandmother, being an example out of many different complex and meaningful stimuli that would have their own cells in the brain. Although this would be a fragile and wasteful encoding, researchers reported grandmother cell like neurons in the medial temporal cortex (Quiroga et al., 2005). Responses of these cells were later specified as sparse and not grandmother-like representations (Quiroga et al., 2008). This novel interpretation of the experimental findings harmonizes with our sparse encoding concepts.

2.6.4 'Mirror neurons'

Certain neurons in primate brains respond, e.g., to hands no matter if the hand belongs to the observer or not (Gallese et al., 1996; Rizzolatti et al., 1996). Such neurons have been termed 'mirror neurons'.

'Mirror neurons' may emerge in a component extracting system, since hands, alike to other body parts are decomposable, so they are Type 1 CFs. In particular, compression is better if a specific manipulation episode can be remembered independently from who, where and when did it. A person independent representation is more useful due to the compositional nature of CFs. In turn, component extraction, learning, and manipulation offers reasonable explanation for the presence of 'mirror neurons' and impaired component formation may give rise to 'mirror neuron' related impairments. Such impairments have been found (Ramachandran and Oberman, 2009), see more details later.

2.6.5 Predictive Autoencoders

From the point of view of behavior optimization, a serious bottleneck is that at any time instant only a small part of the world is observed, and partial observation can corrupt decision making (Krishnamurthy, 2016). Representations in predictive autoencoders minimize this problem and bring estimations about the uncertainties of future outcomes. Theoretical efforts and computational studies demonstrate the efficiency of this approach (Milacski et al., 2019a, Milacski et al., 2019b).

Now, we turn to *special features found in autism*.

2.6.6 Noisy Brain and Excitation/Inhibition Ratio Imbalance

Learning is harder in noise since either the noise has to be filtered out, or the noise will also be learned. Filtering, however, is to be learned and part of the relevant information may be eliminated if filtering is imprecise. At the level of cognition, noise may affect attention, may give rise to hyperactivity, may restrict behavioral repertoire to avoid corrupted behavior, may give rise to a high variability of responses and so on. It also depends on the environment and the internal reward system.

Rubenstein and Merzenich (2003) suggested that ASD is the result of noise in the brain. Markram et al. (2007) proposed that hyperfunctionalities in reactivity, plasticity, perception, attention, and memory become debilitating in ASD, causing social and environmental withdrawal and locking the individual into a small repertoire of proven routines. In other words, the brain is not noisy, but it is highly responsive, and strong influences are dampened by restricted behaviors. The investigators studied the valproic acid rat model; valproic acid is an anticonvulsant and mood stabilizing drug that has been used for epilepsy and schizophrenia. Based on their studies, they propose that excitations are too high and have behavioral consequences.

Davis and Plaisted-Grant (2015) argue that ASD symptoms reflect too little instead of too much neural noise. They argue that (i) the stochastic resonance observed in single unit recordings can take advantage of additive noise and may give rise to improved detection and discrimination thresholds; (ii) noise facilitates transitions between observations, but such transitions can be slow or even missing in binocular rivalry, as reported for ASD (Robertson et al. 2013). Furthermore, (iii) given similar inputs belonging to the same category, generalization between the inputs may become easier

if the noise level is higher, a well-known strategy used in nonlinear denoising autoencoders (Vincent et al., 2008). However, generalization processes seem inefficient in autism (Plaisted, 2000, 2001).

David et al. (2016) consider not the strengths but the variability of cortical oscillation patterns. They note that searches for abnormal power spectra provide inconsistent results in autism. They emphasize that trial-to-trial variability in cortical oscillations form operational noise in neuronal networks that should consistently communicate between remote areas, and such variations have been found in ASD.

Dickinson et al. (2016) reviewed the literature on excitation-inhibition balance in ASD and found that imbalances are essential and there is supporting evidence for such changes. However, taken together, the evidence justifies neither a net increase in excitation nor a net increase in inhibition in autism.

Recent studies of infants with ASD led to intriguing novel findings on cascades of network efficiencies (Lewis et al., 2017). Network inefficiencies were found in infants at high risk of later ASD before symptom consolidation. Inefficiencies – detected by MRI seed-based tractography measures of connection length and strength – were first apparent in low-level sensory processing as early as the age of 6 months, but only in short-range cortico-cortical connectivity. Inefficiency then spread to higher-level processing, and ASD symptoms appeared. Symptom severity can be predicted by the inefficiencies measured much earlier in low-level processing. Lewis et al. suggest that children with ASD may suffer from diminished synaptic pruning during early development.

In sum, endogenous noise and excitation-inhibition imbalance in the context of ASD are controversial but seem to be present in different cohorts measured in different ways and with different extents of ASD symptoms (Dickinson et al., 2016).

2.6.7 Cortical Hyperexpansion

The genetic background of brain overgrowth is intensively studied as noted earlier in the mTOR Pathway section of this paper. Piven et al. (2017) studied high-risk siblings later diagnosed with autism early in their first year and performed neuroimaging studies beyond their first year. They found cortical area hyperexpansion in the first year. In the second year, brain volume overgrowth followed and seemed to be associated with the emergence of social deficits. In this respect, they also note the heterogeneity and the 'reproducibility crisis' in autism. The relevance of these findings is shown in (Hazlett et al., 2017b): within high-risk siblings, information about brain surface area can predict a positive diagnosis of autism with predictive and sensitivity values of 81% and 88%, respectively.

Larger surface areas of the same volume involve different gyrification patterns, as have been reported in the literature for the left pre- and post-central gyrus (Ecker et al., 2016). Increased gyrification seems to enable an increase in the number of short-distance connections reviewed by Courchesne and Pierce (2005).

2.6.8 Properties of Cells of Sparse Representation in Different Brain Areas and Mammalian Species

Cartesian Factors correspond to the hierarchy of unique high order correlations, that at some level correspond to the concept of the 'grandmother cell'. Taking the grandmother as an example, she would be one of the CFs and any color of the cloth, any style of the cloth, her many poses and other details are represented by other CFs. Such representation is (1) highly compressed, since the number of the CFs can be much smaller than the number of the sensors, e.g., the neurons on the retina, whereas (2) only a few of the CFs may be sufficient for recognizing her in a given situation. Another example is Halle Berry shown on photos from different views, with and without sunglasses, dressed typical or in the Catwoman dress, shown in line drawing, and by the letters of her name in a uniform black background. The same cell had negligible responses to everything else researchers presented (Quiroga et al., 2005).

Cells that respond vigorously to a single person were also found in the medial temporal lobe (MTL), but this feature is not unique to the MTL. In the anterior hippocampus, cells responded in a similar fashion to other celebrities and politicians. Another cell – out of the million cells in that lobe – showed specific responses to the Sidney Opera House and the Baha'i Temple, also when shown by letters only, but baseline responses were found for the Eiffel Tower, the Leaning Tower of Pisa, and so on. However, there were single cells that responded e.g., to more than one actor from the TV series 'Friends' or to more than one towers from Europe.

The phenomenon is not restricted to human subjects. There are 'place cells' in the rodent hippocampus that respond to specific local regions of space independently to light conditions, the head direction of the animal. see, e.g., (Andersen et al. 2006). Response fields of these cells can be distorted by distorting the environment, by modifying distant and proximal cues and are totally abolished but get reorganized later when the animal enters a novel environment. These cells respond sparsely; there are place cells with overlapping receptive fields.

The phenomenon is somewhat similar in the primary visual cortex. Here, the so-called simple cells respond optimally to moving rectangular bars of specific orientation, i.e., to a specific higher order correlation pattern in natural scenes (DeAngelis et al., 1993). Such higher order correlations are relatively frequent in natural scenes as opposed to random noise, but the responses of these units are sparse in time and there are cells with overlapping spatial receptive fields. Furthermore, responses are influenced by the environment around the receptive field of the cell, i.e., by the context of the excitation (Chen, Dan, and Li, 2005).

We summarize the properties of the sparse representation:
- (a) Cells respond sparsely.
- (b) They respond to high order correlations.
- (c) Representation may be close to the grandmother cell concept, with the following note: it is not a single cell, which is responding to the high order correlation and responses are not restricted to single entities.
- (d) Responses are strong for diverse representations of the entity. Considering celebrities, for example, responses are view and dress invariant, they can be invariant to the style of the image (photo or line drawing) and strong responses appear for the name alone.

In turn, simple cells, 'grandmother cell-like neurons', and place cells respond to a non-linear subspace of the enormous sensory input space and (a) the subspace has many details about the entity represented by the cell, (b) a small portion of the subspace is sufficient for the cell to provide considerable output and this response is highly specific to that entity. The same input is enough for the subject to recall other parts embraced by the subspace (Quiroga et al., 2005). In turn, pattern completion ability is another specific feature of these cells.

2.6.9 Differences in Axial Diffusivity, Myelination and Synaptic Homeostasis

According to Buzsáki et al. (2013), axon size and myelination seem to be the most important factors for the scaling of network oscillations because they determine the conduction velocity of neurons. The slower the conduction velocity, the larger the delay to be compensated by predictive models to be learned and, in turn, the slower (harder) the learning may become.

 Ecker et al., (2016) found that increased gyrification is accompanied by atypical neural axial sprouting, which is most pronounced in axons traveling close to the cortical sheet. They report that enhanced gyrification correlates with increased axial diffusivity in general, and the relationship may be causal in either direction. Increased axonal sprouting may give rise to erroneous associations and that may also increase irrelevant incoming information, i.e., the noise level.

2.6.10   Connectivity

Synaptogenesis is an additional vital matter. Changes in spine morphology can be vital in forming large networks, and troubled synaptic connections are considered the main underlying reason for autism by many researchers (Toro et al., 2010; Penzes et al., 2011). According to the data, increased spine density gives rise to decreases in cognitive functioning in ASD, supporting the view that ASD can be characterized by denser connectivities locally and hypoconnectivity globally (J. R. Hughes, 2007). Synaptic genes are among the major ASD candidate genes, as described before.

2.6.11   EEG abnormalities

Brain wave abnormalities should affect component forming as well as the cognitive manipulation of components. Consistently, there are EEG abnormalities in autism and the abnormalities correlate with associated phenotypic features (Nicotera et al., 2019).

2.6.12   Consolidation of Emotional Components

Emotional conditioning have been dissociated from declarative memory by Bechara et al. (1995). Based on their results the amygdala is necessary for emotional conditioning, while the hippocampus is required for forming declarative memories.

The amygdala theory of autism appeared early (S. Baron-Cohen et al., 2000): the amygdala was proposed to be one of several neural regions that are abnormal in autism. The theory is based on the general agreement that emotions are processed by the amygdala and it is supported by the anatomy: The ventral hippocampus projects directly to the basolateral amygdala and the central amygdala, and connections are reciprocal (Pitkänen et al., 2000),

a) fear can be switched on and off in distinct circuits between the amygdala, hippocampus and the medial prefrontal cortex by selective activation of specific neuronal circuits (Herry et al., 2008), and

b) multiple parallel pathways exist between the amygdala and the hippocampus. One pathway encodes the context-dependent retrieval of cued fear memories. Another pathway is concerned with fear behavior in a context dependent manner (Xu et al., 2016).

In turn, fear is an internal component that characterizes both episodic and semantic memories. Experimental studies on the consolidation of emotional components have been published (Girardeau et al., 2017). They found that reactivations of memory traces in the basolateral amygdala peaked during hippocampal SPW ripples, being in line with other consolidation patterns.

Recent studies show that in comparison with controls, amygdala volume is greater in ASD (Gibbard et al., 2018). In addition, it was found that higher endogenous oxytocin levels correlate with weaker functional coupling between amygdala and hippocampal regions, in adults with ASD, suggesting weaker attachment scores and that was also found experimentally (Alaerts et al., 2019).

3   The Autism Palette: A Framework for Autism Research

The two main requirements of good social interactions are (i) the motivation to be social and (ii) the capability for solving social problems by selecting the hidden components, i.e., the relevant CFs out of the many possible ones. We start with autism theories, their advantages and disadvantages from the point of view of our model and finish the section by reviewing comorbidities with ASD. We argue that comorbidities are the clinical consequences of the diverse changes in the neural system and that their diversity supports our model.

3.1   Comparing Our Complexity Based Arguments to Cognitive Theories of Autism

There are many cognition-based models of autism supported by both information theory and experiments.

3.1.1 Cognitive Impairments

Cantio et al. (2016) studied cognitive-level symptoms and searched for a universal pattern of cognitive impairments in ASD. They found that two such impairments - (i) impaired theory of mind, i.e., theorizing about the hidden mental states of other people (Ozonoff, 1991), and (ii) impaired executive function, manifested as repetitive and stereotyped behaviors, among other characteristics – predict autism at a rate of up to 75% (50% being a random association). Tests on embedded figures (Spreen and Benton, 1969) showed that local processing bias has non-significant contribution with close to normal distributions of similar variances. Their results support the idea that the social behavior impairments in autistic individuals may arise from potentially different causes. Cognitive impairments, however, provide no explanation for the IQ distribution in ASD and, in particular, for the relatively high number of autistic individuals with above average IQ.

3.1.2 Weak Central Coherence Theory

Weak Central Coherence (WCC) theory is an early insightful model for autism. Happé and Frith (1994) put forth the idea that autistic behavior is the result of impairments in extracting global form and meaning. Later, the model was modified (Happé and Frith, 2006) to say that problems might arise from the superiority of local processing, which is a bias in the processing strategy, and weak

coherence may not be the cause but a symptom of autistic behavior. Robertson and Baron-Cohen (2017) object to the theory on the grounds that it is a top-down mechanism and that bias in the cognitive strategy can hardly explain low-level sensory processing features. However, if top-down mechanisms are supported by distributed sparse representations in an autoencoder and if the two types of Cartesian Factors are imperfect, then central coherence can be weak due to the lack of some CFs, or the synchronization process that serves coherence, or both.

3.1.3 Bayesian Prior Theory

Pellicano and Burr (Pellicano and Burr, 2012) suggested the use of Bayesian models to understand autistic information processing. According to them, differences lie in the perceptual mechanisms; namely, people with ASD have 'hypo-priors' that give rise to unique, highly precise perceptual experiences. Based on this assumption, Pellicano and Burr claim that many autistic characteristics, from sensory processing to non-social impairments stem from differences in Bayesian prediction processes.

 A recent work by Palmer et al. (2017) summarizes the proposal that in autism, sensory information has larger weighting than in normal people. They argue that the balance between perception and action may be the characteristic difference between people with autism and normal people – in both social and non-social behaviors.

There is little doubt that Bayesian inference is difficult to discount, and it seems that this strategy is applied by the brain, see, e.g., (Berkes et al., 2011; Lee and Mumford, 2003; Robertson and Baron-Cohen, 2017) and the cited references therein. On the other hand, the world – apart e.g., from partial observations – is close to deterministic and actions with deterministic outcomes are both possible and desired and the Bayesian account may be limited in this respect: Experimental findings on endogenous and exogenously modulated binocular rivalry (Brascamp et al., 2018) seem to contradict simple Bayesian principles, since a Bayesian observer would always pick the higher-probability interpretation. The switching phenomenon, i.e., that perception switches from one potential interpretation to the other, however, indicates that simple Bayesian models are incomplete. Still, the fact that binocular rivalry is slowed down in autism seems consistent with the Bayesian observer assumption.

3.2   'Mirror Neuron Theory'

Since the discovery of mirror neurons (Gallese et al., 1996; Rizzolatti et al., 1996), that react similarly for goal-oriented self-motions and for similar motions of others and thus allow estimations of the intentions of others, scientists have considered that the mirror neuron system may be impaired or possibly dysfunctional in autism, see, e.g., (Williams et al., 2001; Rizzolatti and Fabbri-Destro 2010; Hillus et al. 2019, and the cited references therein). It was thought that in typical cases, mirror neurons can provide supervisory information for training and for copying behavioral templates.

Experiments support this assumption to some extent: μ waves are typically blocked or reduced during voluntary muscle movement, e.g., when opening or closing the hands, regardless of whether the subject makes the movement or observes someone else making it, but they were not blocked even in high-functioning autistic children when they monitored someone else's muscle movement (Oberman et al., 2005).

These findings on μ waves used to be considered compelling (but, in our view, only implicit) evidence for problems with the mirror neuron system (Ramachandran and Oberman, 2006). Indeed, contradicting evidence has been found. For example, (Bird et al., 2007; Sowden et al., 2016) showed that automatic imitation is intact in ASD. Furthermore, upon separating automatic imitation from spatial compatibility effects (i.e., separating responses on the same side and on the opposite side), there was no relationship between spatial compatibility and autism symptom severity, meaning that individuals with ASD exhibited increased (and not decreased) imitations. The phenomenon is called hyperimitation (Bird et al., 2007; Spengler et al., 2010; Sowden et al., 2016; Deschrijver et al., 2017). These findings are supported by evidence that individuals with ASD frequently engage in strong imitative behavior, such as echolalia and echopraxia.

We note that component learning should support the development of hand representations independent from the owner of the hand to decrease the curse of dimensionality. However, imitation does not require this separation and lacking this piece of information imitation may become similar to repetition, and that could account for the μ wave-related findings.

3.3   Genetic Underpinnings

3.3.1 Comorbidities Depend on the Type of Genetic Variants

Severity of ASD is highly dependent on the manifestation of comorbid conditions as macrocephaly, epilepsy, schizophrenia and ID. These comorbidities are common in ASD and through impairing intelligence they are influencing ASD ethology. However, a minor, but notable portion of autistic individuals does not have any severe comorbidity. Concerning IQ, about one quarter of the ASD cases have severe ID, while half of the cases have normal or above average IQ (Christensen et al., 2018).

We hypothesize that ASD risk increasing genetic variants have small to medium influences, but on many vulnerability pathways, creating a disease-causing combination of impairments (van de Lagemaat and Grant, 2010). To study the relation between genetics and diseases we should separate the types of the causative genetic variants. The ratio of rare, disruptive alleles is higher in severe, than in mild ID, while inherited common variants are more strongly connected to mild, than to severe ID (Kurki et al., 2019). Moreover, ASD related, rare, disruptive variants are purified by negative selection, but common variants are under positive selection (Polimanti and Gelernter, 2017). Interestingly, these latter variants are positively associated with intelligence (Clarke et al., 2016). These results suggest that although common ASD variants can be evolutionarily beneficial, cognition decreasing, rare variants are under negative selection. A strong impairment of component formation would cause a very low level of cognitive abilities in general and that could be compensated only in less complex tasks by increased interest for the components necessary for those tasks. Strong de novos can cause severe IDs that give rise to autistic-like symptoms even without impairments in social rewarding. Such cases may have more severe comorbidities, such as, ID and/or epilepsy. They can be the consequences of epilepsy, abnormal brain development, severely damaged synaptogenesis and the highly penetrant variants behind these neurological problems, among other ones. Severe damages in the *mTOR* pathway or loss of *SHANK3* demonstrates these cases.

3.3.2 Transmission of Causative Variants and the Female Protective Model

As we showed in the previous paragraph, the type of causing genetic variants have a stratificational value on clinical phenotypes. In case of ASD with severe intellectual disability (ID), de novo variants are more frequent, while in high-functioning ASD cases, inherited common variants can be observed. These inherited variants and family history of psychiatric disorders are positively correlated with IQ. It is striking that this stronger familial influence is observable only in high functioning male patients (Robinson et al., 2014). In female cases, comorbidities, de novo variants and rare inherited variants with loss-of-function effects are enriched. Therefore, lower-functioning cases and female cases have a stronger influence from sporadic genetic variants, while high-functioning male cases are influenced more likely by inherited common variants (Robinson et al., 2014).

A recent study tested if shared variants contribute to the disorder by using a standard measure of genetic relation. They compared ASD individuals with unrelated discordant siblings, i.e., unrelated probands and their nonaffected siblings. According to the genetic metric, affected individuals were more similar to the affected than to the unaffected member of the unrelated sibling pair (Ye et al., 2017) as expected. However, common variants and less common, non-coding regulatory variants of dosage sensitive ASD-related genes inherited more likely from the father (Brandler et al., 2018), while rare, disruptive, coding variants of these genes are mostly transmitted from the mother (Iossifov et al., 2015). A possible explanation to this is that variants with moderate effects can be balanced both in man and woman, while variants with larger effects are carried more likely by the mother because of the "female protective model" (Zhao et al., 2007).

Genetic variants of OXT receptor provide further information from multiple points of view. Beyond their effect on prosocial behavior, they show sex-dependent effect on ASD etiology (Dumais and Veenema, 2016). Results suggest that social interaction is more rewarding for women than for men (Borland et al., 2019; Feng et al., 2015; Soutschek et al., 2017). Internal motivation for being in social situations helps one to gain skills in social interaction and practicing. Social rewarding is stronger in females that may be one of the reasons of the "female protective model". Higher burden of mutations and increased ratio of ID in female ASD cases may be due to the social reward system that compensates or possibly overcomes the effect of mild impairments of cognition related social skills.

3.4   Clinical Consequences: Comorbidities

3.4.1 Epilepsy

Epilepsy is a neurological condition. It is often comorbid with other neurologic and psychiatric disorders. Epilepsy shows clinical overlap with ASD and has many other comorbid profiles. The prevalences of epilepsy in males and females are approximately 18% and 34%, respectively (Amiet et al., 2008).

Epilepsy is considered a network problem (Kanner, 1943). As such, it is fragile in many ways, including network centrality (Balardin et al., 2015). Seizures may be due to high excitation/inhibition ratios, which are also one of the main theoretical routes proposed in autism models (Rubenstein and Merzenich, 2003). ASD and epileptic encephalopathy seem to have many common genetic causes (Srivastava and Sahin, 2017).

Cognitive impairment is the most common outcome of epilepsy; epilepsy can cause considerable harm to the developing brain (Holmes, 2016): epilepsy gives rise to morphological and physiological changes, modified synaptogenesis and altered excitatory and/or inhibitory balance, which destroy both network structure and dynamics and increase the severity of the component formation impairment.

Mesial temporal lobe is the cornerstone of component-formation and memory consolidation. In turn, it is a falsifying issue what happens in case of mesial temporal lobe epilepsy that corrupts this key structure. According to (Okruszek et al., 2017), such epilepsy gives rise to social cognitive deficits, including significant mentalizing deficits, supporting our theory. Furthermore, Gelinas et al. (2016) found that spontaneous hippocampal interictal epileptiform discharges correlate with impaired memory consolidation offering a straightforward connection between epilepsy and the impairment of component formation and thus the comorbidity of ASD and epilepsy.

3.4.2 Schizophrenia

Definitions of ASD and schizophrenia have changed over time. ASD, which was first described by Kanner (1943) and Asperger (1944) used to be considered as an early version of schizophrenia (Bender, 1947) or a central feature of schizophrenia (Bleuler, 1950). The opinion that they are different disorders started later (Rutter, 1972), since ASD starts during childhood and is characterized by deficits in social interaction and communication, whereas schizophrenia typically has a later onset and is characterized by psychotic symptoms.

In addition to behavioral phenotypes, there are genetic links between ASD and schizophrenia; see, e.g., (Pina-Camacho et al., 2016) and the references therein. We list a few similarities and dissimilarities in binding and low complexity component formation in epilepsy, schizophrenia and in ASD, since these may influence the learning of Cartesian Factors.

Binding of sensory information between different modalities helps the fusion of information and helps pattern completion when part of the fused information is missing. Problems with bindings have been observed both in schizophrenia and ASD, e.g., in pairing audio and visual signals and in binding interoceptive signals (Noel et al., 2018). Binding requires precise temporal windows, but experiments with ASD patients show expanded audiovisual temporal binding windows and completely diminished temporal acuity for perceiving cardiovisual (interoceptive to exteroceptive) information (Noel et al., 2018). However, it has been questioned, if corruption of binding information sources has the same causes in ASD and in schizophrenia (Noel et al., 2017).

Errors in binding give rise to errors in predictive model learning. Take self-tickling as an example. Insensitivity to self-tickling should not be surprising given a precise model of the self, and indeed, typical individuals cannot tickle themselves. This remains the case even if bodies are swapped in the body transfer illusion (Van Doorn et al., 2014). However, self-produced touch results in more ticklish perceptions in individuals with Asperger's syndrome than in normal subjects (Blakemore et al., 2006), and self-tickling is particularly successful for individuals with pronounced schizotypal traits (Lemaitre et al., 2016).

A more complex pattern appears for the so-called 'rubber hand illusion'. Phantom limbs have been studied extensively over the years (see, e.g., Pirowska et al., 2014 and the references cited). Related findings concern the fast malleability of body representations into the environment, see, e.g.,

Tsakiris et al. (2011). Experimenting with such fast malleability, as in the rubber hand illusion, Noel et al. (2017) found that patients with schizophrenia and ASD behave very differently. Schizophrenia patients had a weak or variable bodily boundary between the self and the environment, whereas ASD patients had a sharp boundary. We argue that the complexity of the separation of the self from the environment corresponds to the separation of self-controlled components from the rest of the world and thus, such separation seems relatively simple among the problems of component learning. In turn, as mentioned before for the case of IQ tests, much less impairment is expected in low-complexity task compared with higher-complexity tasks for ASD patients. Furthermore, since learning is more focused on the self than on partners in ASD, learned boundaries may be more rigid for individuals with ASD.

3.4.3 Synesthesia

Baron-Cohen et al. (2013) report that the incidence of synesthesia, in which a sensation in one modality involves perception in another one, is approximately three times higher in autistic adults than in normal subjects. (Hughes et al., 2017) found that synesthesia in autism is linked to savant skills. These results are further supported by Ward et al., (2018), who showed that synesthetes have enhanced perception and attention and exhibit autistic-like impairments, too. It has been found that axonal connections between V4, which is involved in color processing, and the so-called 'grapheme area' are denser in synesthetes than in controls (Rouw and Scholte, 2007). Both of these areas are in the fusiform gyrus, and some portions can be adjacent. Other findings support a cross-activation model (Ramachandran and Seckel, 2015) between these areas, and some of these cross-activations seem to be preconscious. In turn, the comorbidity of autism and synesthesia seems to arise from network effects, e.g., from increased axial diffusivity, which can be fostered by enhanced perception and attention to either colors or graphemes.

3.5   Combining the arguments

We posed the following question: How come that (a) many subsets of bountiful discrepancies can give rise to a single, although very colorful leading symptom, namely, the impairment in social behaviors and (b) discrepancies may give rise to the leading symptom no matter if some components of those discrepancies are weaker or stronger than usual?

We have argued that
- (a) the searches for adequate behavioral responses in social interactions are to be learned and that learning problem has a high number of spatio-temporal variables that should be deduced from the huge space of sensory information. If so, then there are many potential causes for autism and a few of them may be sufficient to impair social interaction. The set of *potential disfunctions forms the autism spectrum,* whereas the *combinations of elements of the set form* the *autism palette*. The more the number of causes, the less the impairment of the actual components may be that can sum up to have an impact on social interactions.
- (b) genetics reinforces our arguments since many different genetic causes contribute to autism and they act along different pathways and in diverse ways. Genetics thus shows a large set of potential causes and GWASes indicate that ASD individuals have diverse subsets of a much larger set uncovered by GWASes.
- (c) impairments combined in autism include (a) component formation and in (b) reward system for social interaction. This is in line with findings of Warrier et al. (2019) on the genetic dissociability of social and non-social ("systemizing") traits of ASD that translates to our model as follows: social symptoms are consequences of impaired component formation,

while systemizing is the outcome of weakened reinforcement and altered motivation. In addition, a stronger internal reward, i.e., motivation for social interaction may support the female protective model.

We note finally that increasing the frequency of social interaction together with suitable immediate rewards may help the condition, whereas uncovering the combination of the individual causes may be necessary to find efficient specific cognitive therapy and/or medication.

4   Conflict of Interest

5   Author Contributions

Á.F. contributed the most to genetics, A.L. to neuroscience related information and computational principles, L.S. harmonized this multi-disciplinary writing and provided detailed information about autism related questions.

6   Funding

7   Contribution to the Field Statement

This paper outlines a theory of ASD that posits autism is caused by deficits in component-based cognition and the internal learning reinforcement machinery. Specifically, we suggest complex cognition such as social cognition is supported by an autoencoder that supports concept formation, and that this mechanism is disrupted in autism. The paper argues that internal models analyze hidden attributes of partners in complex social interactions and have predictive power that considers emotion and intention together within the spatio-temporal context. The learning of such models is the central challenge for intelligence and model quality depends on discerning relevant variables as well as inner motivation to practice and learn social emotional signs. In turn, corrupted extraction of the relevant emotional and other variables as well as limited social interest increase the strength of autism symptoms. On the other hand, stronger social motivation and better emotion estimation may alleviate the autistic symptoms and offer explanation for the female protective model. Such considerations are connected to diverse psychiatric disorders, etiological theories, and recent genetic findings. The paper suggests that precise measurement of the "color set" of the autistic individual followed by adjusted personalized treatment methods need to be developed.

8   References

Alaerts, K., Bernaerts, S., Vanaudenaerde, B., Daniels, N., and Wenderoth, N. (2019). Amygdala–Hippocampal Connectivity Is Associated With Endogenous Levels of Oxytocin and Can Be Altered by Exogenously Administered Oxytocin in Adults With Autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(7), 655–663. https://doi.org/10.1016/j.bpsc.2019.01.008

Amiet, C., Gourfinkel-An, I., Bouzamondo, A., Tordjman, S., Baulac, M., Lechat, P., et al. (2008). Epilepsy in Autism is Associated with Intellectual Disability and Gender: Evidence from a Meta-Analysis. *Biological Psychiatry*, *64*(7), 577–582. https://doi.org/10.1016/j.biopsych.2008.04.030

Andersen, P., Morris, R., Amaral, D., Bliss, T., O'Keefe, J., and Keefe, J. O. (2006). *The Hippocampus Book*. https://doi.org/10.1093/acprof:oso/9780195100273.001.0001

Angioni, L., Cocco, C., Ferri, G. L., Argiolas, A., Melis, M. R., and Sanna, F. (2016). Involvement of nigral oxytocin in locomotor activity: A behavioral, immunohistochemical and lesion study in male rats. *Hormones and Behavior*, *83*, 23–38. https://doi.org/10.1016/j.yhbeh.2016.05.012

Asperger, H. (1944). Die „Autistischen Psychopathen" im Kindesalter. *European Archives of Psychiatry and Clinical Neuroscience*, *117*(1), 76–136.

Balardin, J. B., Comfort, W. E., Daly, E., Murphy, C., Andrews, D., Murphy, D. G. M., et al. (2015). Decreased centrality of cortical volume covariance networks in autism spectrum disorders. *Journal of Psychiatric Research*, *69*, 142–149. https://doi.org/10.1016/j.jpsychires.2015.08.003

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., et al. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705), 429-433.

Baio, J. et al. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. Morbidity and Mortality Weekly Report. Surveillance summaries, 63(2), 1-21.

Baraniuk, R. G., Cevher, V., Duarte, M. F., and Hegde, C. (2010). Model-Based Compressive Sensing. IEEE *Transactions on Information Theory,* 56(4), 1982-2001.

Bariselli, S., Hörnberg, H., Prévost-Solié, C., Musardo, S., Hatstatt-Burklé, L., Scheiffele, P., and Bellone, C. (2018). Role of VTA dopamine neurons and neuroligin 3 in sociability traits related to nonfamiliar conspecific interaction. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05382-3

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., and Williams, S. C. R. (2000). The amygdala theory of autism. *Neuroscience and Biobehavioral Reviews*, *24*(3), 355–364. https://doi.org/10.1016/S0149-7634(00)00011-7

Baron-Cohen, Simon, Johnson, D., Asher, J., Wheelwright, S., Fisher, S. E., Gregersen, P. K., and Allison, C. (2013). Is synaesthesia more common in autism? *Molecular Autism*, *4*(1), 40. https://doi.org/10.1186/2040-2392-4-40

Bateup, H. S., Johnson, C. A., Denefrio, C. L., Saulnier, J. L., Kornacker, K., and Sabatini, B. L. (2013). Excitatory/Inhibitory Synaptic Imbalance Leads to Hippocampal Hyperexcitability in Mouse Models of Tuberous Sclerosis. *Neuron*, *78*(3), 510–522. https://doi.org/10.1016/j.neuron.2013.03.017

Beauchaine, T. P., and Constantino, J. N. (2017). Redefining the endophenotype concept to accommodate transdiagnostic vulnerabilities and etiological complexity. *Biomarkers in Medicine*, *11*(9), 769–780. https://doi.org/10.2217/bmm-2017-0002

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269*(5227), 1115–1118.

Bender, L. (1947). Childhood schizophrenia: Clinical study of one hundred schizophrenic children. *American Journal of Orthopsychiatry, 17*(1), 40.

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science, 331*(6013), 83–87. https://doi.org/10.1109/TAP.2015.2408364

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*(2), 115.

Bird, G., Leighton, J., Press, C., and Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences*, 274(1628), 3027-3031.

Blakemore, S. J., Tavassoli, T., Calò, S., Thomas, R. M., Catmur, C., Frith, U., and Haggard, P. (2006). Tactile sensitivity in Asperger syndrome. *Brain and Cognition, 61*(1), 5–13. https://doi.org/10.1016/j.bandc.2005.12.013

Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias.* Translated by Bleuler E. and Zinkin J. (1950) New York, NY: International Universities Press.

Bonaglia, M. C., Giorda, R., Borgatti, R., Felisari, G., Gagliardi, C., Selicorni, A., and Zuffardi, O. (2001). Disruption of the ProSAP2 Gene in at(12;22)(q24.1;q13.3) Is Associated with the 22q13.3 Deletion Syndrome. *The American Journal of Human Genetics* (Vol. 69). https://doi.org/10.1086/321293

Borland, J. M., Aiani, L. M., Norvelle, A., Grantham, K. N., O'Laughlin, K., Terranova, J. I., et al.(2019). Sex-dependent regulation of social reward by oxytocin receptors in the ventral tegmental area. *Neuropsychopharmacology, 44*(4), 785–792. https://doi.org/10.1038/s41386-018-0262-y

Born, J., and Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research, 76*(2), 192–203. https://doi.org/10.1007/s00426-011-0335-6

Bourgeron, T. (2009). A synaptic trek to autism. *Current Opinion in Neurobiology, 19*(2), 231–234. https://doi.org/10.1016/j.conb.2009.06.003

Boyce, R., Glasgow, S. D., Williams, S., and Adamantidis, A. (2016). Sleep research: Causal evidence for the role of REM sleep theta rhythm in contextual memory consolidation. *Science, 352*(6287), 812–816. https://doi.org/10.1126/science.aad5252

Brandler, W. M., Antaki, D., Gujral, M., Kleiber, M. L., Whitney, J., Maile, M. S., et al.(2018). Paternally inherited cis-regulatory structural variants are associated with autism. *Science, 360*(6386), 327–331. https://doi.org/10.1126/science.aan2261

Brascamp, J., Sterzer, P., Blake, R., and Knapen, T. (2018). Multistable perception and the role of the frontoparietal cortex in perceptual inference. *Annual Review of Psychology, 69*(1), 77–103. https://doi.org/10.1146/annurev-psych-010417-085944

Bressan, P., Mingolla, E., Spillmann, L., and Watanabe, T. (1997). Neon color spreading: a review. *Perception*, 26(11), 1353-1366.

Bricolo, E., Poggio, T., and Logothetis, N. K. (1997). 3D object recognition: A model of view-tuned neurons. *Advances in Neural Information Processing Systems* (pp. 41-47).

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Burrows, E. L., Laskaris, L., Koyama, L., Churilov, L., Bornstein, J. C., Hill-Yardin, E. L., and Hannan, A. J. (2015). A neuroligin-3 mutation implicated in autism causes abnormal aggression and increases repetitive behavior in mice. *Molecular Autism*, *6*(1), 1–11. https://doi.org/10.1186/s13229-015-0055-7

Buzsáki, G. (1989). Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience*, Vol. 31, pp. 551–570. https://doi.org/10.1016/0306-4522(89)90423-5

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. *Neuron*, *80*(3), 751–764. https://doi.org/10.1016/j.neuron.2013.10.002

Cantio, C., Jepsen, J. R. M., Madsen, G. F., Bilenberg, N., and White, S. J. (2016). Exploring 'The autisms' at a cognitive level. *Autism Research*, *9*(12), 1328–1339. https://doi.org/10.1002/aur.1630

Chen, G., Dan, Y., and Li, C. Y. (2005). Stimulation of non-classical receptive field enhances orientation selectivity in the cat. *Journal of Physiology*, *564*(1), 233–243. https://doi.org/10.1113/jphysiol.2004.080051

Chih, B., Engelman, H., and Scheiffele, P. (2005). Control of excitatory and inhibitory synapse formation by neuroligins. *Science*, *307*(5713), 1324–1328. https://doi.org/10.1126/science.1107470

Chrisley, R., and Sloman, A. (2016). Functionalism, revisionism, and qualia. *APA Newsletter on Philosophy and Computers*, 16(1), 2-13.

Christensen, D. L., Braun, K. V. N., Baio, J., Bilder, D., Charles, J., Constantino, J. N., et al. (2018). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR. Surveillance Summaries*, *65*(13), 1–23. https://doi.org/10.15585/mmwr.ss6513a1

Chrobak, J. J., Lörincz, A., and Buzsáki, G. (2000). Physiological patterns in the hippocampo entorhinal cortex system. *Hippocampus*, 10(4), 457-465.

Clarke, T. K., Lupton, M. K., Fernandez-Pujals, A. M., Starr, J., Davies, G., Cox, S., et al. (2016). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*, *21*(3), 419–425. https://doi.org/10.1038/mp.2015.12

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., et al. (2015). Heritability of autism spectrum disorder in a UK population-based twin sample. *JAMA Psychiatry*, *72*(5), 415–423. https://doi.org/10.1001/jamapsychiatry.2014.3028

Constantinescu, A. O., O'Reilly, J. X., and Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science* (New York, NY), 352(6292), 1464-1468.

Cotney, J., Muhle, R. A., Sanders, S. J., Liu, L., Willsey, A. J., Niu, W., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nature Communications*, *6*, 1–11. https://doi.org/10.1038/ncomms7404

Courchesne, E., Karns, C. M., Davis, H. R., Ziccardi, R., Carper, R. A., Tigue, Z. D., et al. (2001). Unusual brain growth patterns in early life in patients with autistic disorder: An MRI study. *Neurology*, *57*(2), 245–254. https://doi.org/10.1212/WNL.57.2.245

Courchesne, E., and Pierce, K. (2005). Brain overgrowth in autism during a critical time in development: Implications for frontal pyramidal neuron and interneuron development and connectivity. *International Journal of Developmental Neuroscience*, *23*(2-3 SPEC. ISS.), 153–170. https://doi.org/10.1016/j.ijdevneu.2005.01.003

Crawford, D. C., Acuña, J. M., Sherman, S. L., and Med, G. (2001). FMR1 and the Fragile X Syndrome: Human Genome Epidemiology Review HHS Public Access Author manuscript. *Genet Med*, *3*(5), 359–371.

Cover, T. M., and Thomas, J. A. (2012). *Elements of information theory*. John Wiley and Sons.

Crino, P. B. (2011). MTOR: A pathogenic signaling pathway in developmental brain malformations. *Trends in Molecular Medicine*, *17*(12), 734–742. https://doi.org/10.1016/j.molmed.2011.07.008

Cullen, K. E., and Taube, J. S. (2017). Our sense of direction: progress, controversies and challenges. *Nature Neuroscience*, 20(11), 1465-1473.

David, M. M., Enard, D., Ozturk, A., Daniels, J., Jung, J. Y., Diaz-Beltran, L., and Wall, D. P. (2016). Comorbid analysis of genes associated with autism spectrum disorders reveals differential evolutionary constraints. *PLoS ONE*, *11*(7), 1–15. https://doi.org/10.1371/journal.pone.0157937

David, N., Newen, A., and Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition*, *17*(2), 523–534. https://doi.org/10.1016/j.concog.2008.03.004

Davis, G., and Plaisted-Grant, K. (2015). Low endogenous neural noise in autism. *Autism*, *19*(3), 351–362. https://doi.org/10.1177/1362361314552198

de la Torre-Ubieta, L., Won, H., Stein, J. L., and Geschwind, D. H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nature Medicine*, *22*(4), 345–361. https://doi.org/10.1038/nm.4071

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209–215. https://doi.org/10.1038/nature13772

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, *69*(4), 1091–1117. https://doi.org/10.1152/jn.1993.69.4.1091

Deschrijver, E., Wiersema, J. R., and Brass, M. (2017). The influence of action observation on action execution: Dissociating the contribution of action on perception, perception on action, and resolving conflict. *Cognitive, Affective, and Behavioral Neuroscience*, 17(2), 381-393.

Desolneux, A., Moisan, L., and Morel, J. M. (2004). Gestalt theory and computer vision. *Seeing, Thinking and Knowing* (pp. 71-101). Springer, Dordrecht.

Devlin, B., and Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Current Opinion in Genetics and Development*, *22*(3), 229–237. https://doi.org/10.1016/j.gde.2012.03.002

Dickinson, A., Jones, M., and Milne, E. (2016). Measuring neural excitation and inhibition in autism: Different approaches, different findings and different interpretations. *Brain Research* (Vol. 1648). https://doi.org/10.1016/j.brainres.2016.07.011

Diego, F., and Hamprecht, F. A. (2013). Learning multi-level sparse representations. *Advances in Neural Information Processing Systems (*pp. 818-826).

Diekelmann, S., and Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Domingos, P. (1998). Occam's two razors: The sharp and the blunt. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*. AAAI Press (pp. 37-43)

Dumais, K. M., and Veenema, A. H. (2016). Vasopressin and oxytocin receptor systems in the brain: Sex differences and sex-specific regulation of social behavior. *Frontiers in Neuroendocrinology*, *40*, 1–23. https://doi.org/10.1016/j.yfrne.2015.04.003

Dupre, C., and Yuste, R. (2017). Non-overlapping Neural Networks in Hydra vulgaris. *Current Biology*, *27*(8), 1085–1097. https://doi.org/10.1016/j.cub.2017.02.049

Eccles, J. C. (2013). *The physiology of synapses*. Academic Press.

Ecker, C., Andrews, D., Dell'Acqua, F., Daly, E., Murphy, C., Catani, M., et al. (2016). Relationship between cortical gyrification, white matter connectivity, and autism spectrum disorder. *Cerebral Cortex*, *26*(7), 3297–3309. https://doi.org/10.1093/cercor/bhw098

Efeyan, A., and Sabatini, D. M. (2010). MTOR and cancer: Many loops in one pathway. *Current Opinion in Cell Biology*, *22*(2), 169–176. https://doi.org/10.1016/j.ceb.2009.10.007

Ekman, P. (1982). Methods for measuring facial action. *Handbook of methods in nonverbal behavior research*.

Ernst, M., Zametkin, A. J., Matochik, J. A., Pascualvaca, D., and Cohen, R. M. (1997). Low medial prefrontal dopaminergic activity in autistic children. *Lancet*, *350*(9078), 638. https://doi.org/10.1016/S0140-6736(05)63326-0

Feng, C., Hackett, P. D., DeMarco, A. C., Chen, X., Stair, S., Haroon, E., et al. (2015). Oxytocin and vasopressin effects on the neural response to social cooperation are modulated by sex in humans. *Brain Imaging and Behavior*, *9*(4), 754–764. https://doi.org/10.1007/s11682-014-9333-9

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12), 11-39.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., and Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316. https://doi.org/10.1073/pnas.0706111104

Friesen, A. L., and Domingos, P. M. (2018). Submodular field grammars: representation, inference, and application to image parsing. *Advances in Neural Information Processing Systems* (pp. 4307-4317).

Frith, U., and Happé, F. (1994). Autism: Beyond "theory of mind." *Cognition*, *50*(1–3), 115–132. Retrieved from

Fuccillo, M. V. (2016). Striatal circuits as a common node for autism pathophysiology. *Frontiers in Neuroscience*, *10*(FEB). https://doi.org/10.3389/fnins.2016.00027

Fujii, S., and Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Frontiers in Human Neuroscience*, 8, 777.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119 (Pt 2*(September), 593–609. https://doi.org/10.1093/brain/119.2.593

Gandal, M. J., Haney, J. R., Parikshak, N. N., Leppa, V., Ramaswami, G., Hartl, C., et al. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, *359*(6376), 693–697. https://doi.org/10.1126/science.aad6469

Gao, R., and Penzes, P. (2015). Common Mechanisms of Excitatory and Inhibitory Imbalance in Schizophrenia and Autism Spectrum Disorders. *Current Molecular Medicine*, *15*(2), 146–167. https://doi.org/10.2174/1566524015666150303003028

Garrido Zinn, C., Clairis, N., Silva Cavalcante, L. E., Furini, C. R. G., de Carvalho Myskiw, J., and Izquierdo, I. (2016). Major neurotransmitter systems in dorsal hippocampus and basolateral amygdala control social recognition memory. *Proceedings of the National Academy of Sciences*, *113*(33), E4914–E4919. https://doi.org/10.1073/pnas.1609883113

Gelinas, J. N., Khodagholy, D., Thesen, T., Devinsky, O., and Buzsáki, G. (2016). Interictal epileptiform discharges induce hippocampal–cortical coupling in temporal lobe epilepsy. *Nature Medicine*, *22*(6), 641.

Gibbard, C. R., Ren, J., Skuse, D. H., Clayden, J. D., and Clark, C. A. (2018). Structural connectivity of the amygdala in young adults with autism spectrum disorder. *Human Brain Mapping*, *39*(3), 1270–1282.

Girardeau, G., Inema, I., and Buzsáki, G. (2017). Reactivations of emotional memory in the hippocampus-amygdala system during sleep. *Nature Neuroscience*, *20*(11), 1634–1642. https://doi.org/10.1038/nn.4637

Happé, F., and Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., et al. (2017a). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 348–351. https://doi.org/10.1038/nature21369

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., et al. (2017b). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 348–351. https://doi.org/10.1038/nature21369

Henze, D. A., Wittner, L., and Buzsáki, G. (2002). Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. Nature neuroscience, 5(8), 790

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., and Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature*, *454*(7204), 600–606. https://doi.org/10.1038/nature07166

Heys, J. G., MacLeod, K. M., Moss, C. F., and Hasselmo, M. E. (2013). Bat and rat neurons differ in theta-frequency resonance despite similar coding of space. *Science*, *340*(6130), 363–367. https://doi.org/10.1126/science.1233831

Hillus, J., Moseley, R., Roepke, S., and Mohr, B. (2019). Action semantic deficits are associated with impaired motor skills in autistic adults without intellectual impairment. *Frontiers in Human Neuroscience*, 13:256.

Holmes, G. L. (2016). Effect of Seizures on the Developing Brain and Cognition. *Seminars in Pediatric Neurology*, *23*(2), 120–126. https://doi.org/10.1016/j.spen.2016.05.001

Horev, G., Ellegood, J., Lerch, J. P., Son, Y.-E. E., Muthuswamy, L., Vogel, H., et al. (2011). Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. *Proceedings of the National Academy of Sciences*, *108*(41), 17076–17081. https://doi.org/10.1073/pnas.1114042108

Hughes, J. E. A., Simner, J., Baron-Cohen, S., Treffert, D. A., and Ward, J. (2017). Is Synaesthesia More Prevalent in Autism Spectrum Conditions? Only Where There Is Prodigious Talent. *Multisensory Research*, *30*(3–5), 391–408. https://doi.org/10.1163/22134808-00002558

Hughes, J. R. (2007). Autism: The first firm finding = underconnectivity? *Epilepsy and Behavior*, *11*(1), 20–24. https://doi.org/10.1016/j.yebeh.2007.03.010

Hung, L. W., Neuner, S., Polepalli, J. S., Beier, K. T., Wright, M., Walsh, J. J., et al. (2017). Gating of social reward by oxytocin in the ventral tegmental area. *Science*, *357*(6358), 1406–1411. https://doi.org/10.1126/science.aan4994

Ichtchenko, K., Nguyen, T., and Südhof, T. C. (1996). Structures, alternative splicing, and neurexin binding of multiple neuroligins. *Journal of Biological Chemistry*, *271*(5), 2676–2682. https://doi.org/10.1074/jbc.271.5.2676

Inoki, K., Corradetti, M. N., and Guan, K. L. (2005). Dysregulation of the TSC-mTOR pathway in human disease. *Nature Genetics*, *37*(1), 19–24. https://doi.org/10.1038/ng1494

Ishai, A., and Sagi, D. (1997). Visual imagery facilitates visual perception: Psychophysical evidence. *Journal of Cognitive Neuroscience,* 9(4), 476-489.

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y., et al. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proceedings of the National Academy of Sciences*, *112*(41), E5600–E5607. https://doi.org/10.1073/pnas.1516376112

Jackson, F. (1978). Epiphenomenal qualia. *Philosophical Quarterly*, 15, 249-61.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*(3), 217–250.

Kearns, M., and Koller, D. (1999). Efficient reinforcement learning in factored MDPs. *IJCAI International Joint Conference on Artificial Intelligence*, *2*, 740–747.

Kearns, M., and Singh, S. (2002). Near-optimal reinforcement learning in polynomial time. *Machine Learning*, *49*(2–3), 209–232.

Kelleher, R. J., and Bear, M. F. (2008). The Autistic Neuron: Troubled Translation? *Cell*, *135*(3), 401–406. https://doi.org/10.1016/j.cell.2008.10.017

Kim, H., Lim, C. S., and Kaang, B. K. (2016). Neuronal mechanisms and circuits underlying repetitive behaviors in mouse models of autism spectrum disorder. *Behavioral and Brain Functions*, *12*(1), 1–13. https://doi.org/10.1186/s12993-016-0087-y

Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung*, 31(1), XVIII-XXX.

Krishnamurthy, V. (2016). Partially observed Markov decision processes: From filtering to controlled sensing. Cambridge University Press.

Kruppa, J. A., Gossen, A., Oberwelland Weiß, E., Kohls, G., Großheinrich, N., Cholemkery, H., et al. (2019). Neural modulation of social reinforcement learning by intranasal oxytocin in male adults with high-functioning autism spectrum disorder: a randomized trial. *Neuropsychopharmacology*, *44*(4), 749–756. https://doi.org/10.1038/s41386-018-0258-7

Kurki, M. I., Saarentaus, E., Pietiläinen, O., Gormley, P., Lal, D., Kerminen, S., et al.(2019). Contribution of rare and common variants to intellectual disability in a sub-isolate of Northern Finland. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-018-08262-y

Lainhart, J. E., Piven, J., Wzorek, M., Landa, R., Santangelo, S. L., Coon, H., and Folstein, S. E. (1997). Macrocephaly in children and adults with autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(2), 282–290. https://doi.org/10.1097/00004583-199702000-00019

Larson, M. J., South, M., Krauskopf, E., Clawson, A., and Crowley, M. J. (2011). Feedback and reward processing in high-functioning autism. *Psychiatry Research*, *187*(1–2), 198–203. https://doi.org/10.1016/j.psychres.2010.11.006

Leblond, C. S., Nava, C., Polge, A., Gauthier, J., Huguet, G., Lumbroso, S., et al. (2014). Meta-analysis of SHANK Mutations in Autism Spectrum Disorders: A Gradient of Severity in Cognitive Impairments. *PLoS Genetics*, *10*(9). https://doi.org/10.1371/journal.pgen.1004580

Lee, J. H., Huynh, M., Silhavy, J. L., Kim, S., Dixon-Salazar, T., Heiberg, A., et al. (2012). De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nature Genetics*, *44*(8), 941–945. https://doi.org/10.1038/ng.2329

Lee, T. S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/josaa.20.001434

Lemaitre, A. L., Luyat, M., and Lafargue, G. (2016). Individuals with pronounced schizotypal traits are particularly successful in tickling themselves. *Consciousness and Cognition*, *41*, 64–71. https://doi.org/10.1016/j.concog.2016.02.005

Lewis, J. D., Evans, A. C., Pruett, J. R., Botteron, K. N., McKinstry, R. C., Zwaigenbaum, L., et al. (2017). The Emergence of Network Inefficiencies in Infants With Autism Spectrum Disorder. *Biological Psychiatry*, *82*(3), 176–185. https://doi.org/10.1016/j.biopsych.2017.03.006

Li, N., Lee, B., Liu, R. J., Banasr, M., Dwyer, J. M., Iwata, M., et al. (2010). mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science*, *329*(5994), 959–964. https://doi.org/10.1126/science.1190287

Lipton, J. O., and Sahin, M. (2014). The Neurology of mTOR. *Neuron*, *84*(2), 275–291. https://doi.org/10.1016/j.neuron.2014.09.034

Ljungberg, T., Apicella, P., and Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, *67*(1), 145–163. https://doi.org/10.1152/jn.1992.67.1.145

Lőrincz, A., Szatmáry, B., & Szirtes, G. (2002). The mystery of structure and function of sensory processing areas of the neocortex: A resolution. *Journal of Computational Neuroscience*, 13(3), 187-205. https://doi.org/10.1023/A:1020262214821

Lőrincz, A. (2018). *Colors of Autism Spectrum: A Single Paradigm Explains the Heterogeneity*. https://doi.org/10.20944/preprints201804.0302.v1

Lőrincz, A. (2019). Sparsified and Twisted Residual Autoencoders: Mapping Cartesian Factors to the Entorhinal-Hippocampal Complex. *Biologically Inspired Cognitive Architectures* (pp. 321-332). Springer, Cham.

Lőrincz, A., and Buzsáki, G. (2000). Two phase computational model training long term memories in the entorhinal hippocampal region. *Annals of the New York Academy of Sciences,* 911(1), 83-111.

Lőrincz, A., and Sárkány, A. (2017). Semi-Supervised Learning of Cartesian Factors: A Top-Down Model of the Entorhinal Hippocampal Complex. *Frontiers in Psychology*, 8, 215.

Markram, H., Rinaldi, T., and Markram, K. (2007). The intense world syndrome-an alternative hypothesis for autism. *Frontiers in Neuroscience*, *1*(1), 6. https://doi.org/10.3389/neuro.01.1.1.006.2007

Meyer, M. L., Davachi, L., Ochsner, K. N., and Lieberman, M. D. (2018). Evidence That Default Network Connectivity During Rest Consolidates Social Information. *Cerebral Cortex*, *29*(5), 1910–1920. https://doi.org/10.1093/cercor/bhy071

Milacski, Z. Á., Póczos, B., and Lőrincz, A. (2019a). Group k-Sparse Temporal Convolutional Neural Networks: Unsupervised Pretraining for Video Classification. To appear in: *Proceedings of the International Joint Conference on Neural Networks*.

Milacski, Z. Á., Póczos, B., & Lőrincz, A. (2019b). Differentiable Unrolled Alternating Direction Method of Multipliers for OneNet. To appear in: Proceedings of the British Conference on Machine Vision.

Moser, E. I., Kropff, E., and Moser, M.-B. (2008). Place Cells, Grid Cells, and the Brain's Spatial Representation System. *Annual Review of Neuroscience*, *31*(1), 69–89. https://doi.org/10.1146/annurev.neuro.31.061307.090723

Nguyen, T. M., Schreiner, D., Xiao, L., Traunmüller, L., Bornmann, C., and Scheiffele, P. (2016). An alternative splicing switch shapes neurexin repertoires in principal neurons versus interneurons in the mouse hippocampus. *ELife*, *5*(DECEMBER2016), 1–24. https://doi.org/10.7554/eLife.22757.001

Nicotera, A. G., Hagerman, R. J., Catania, M. V., Buono, S., Di Nuovo, S., Liprino, E. M., et al. (2019). EEG Abnormalities as a Neurophysiological Biomarker of Severity in Autism Spectrum Disorder: A Pilot Cohort Study. *Journal of Autism and Developmental Disorders*, *49*(6), 2337–2347. https://doi.org/10.1007/s10803-019-03908-2

Noel, J. P., Cascio, C. J., Wallace, M. T., and Park, S. (2017). The spatial self in schizophrenia and autism spectrum disorder. *Schizophrenia Research*, *179*, 8–12. https://doi.org/10.1016/j.schres.2016.09.021

Noel, J. P., Lytle, M., Cascio, C., and Wallace, M. T. (2018). Disrupted integration of exteroceptive and interoceptive signaling in autism spectrum disorder. *Autism Research*, *11*(1), 194–205. https://doi.org/10.1002/aur.1880

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., and Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24(2), 190-198.

Okruszek, Ł., Bala, A., Wordecha, M., Jarkiewicz, M., Wysokiński, A., Szczepocka, E., et al. (2017). Social cognition in neuropsychiatric populations: a comparison of theory of mind in schizophrenia and mesial temporal lobe epilepsy. *Scientific Reports*, *7*(1), 484. https://doi.org/10.1038/s41598-017-00565-2

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607-609.

Ozonoff, S., Pennington, B. F., and Rogers, S. J. (1991). Executive function deficits in high functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry,* 32(7), 1081-1105.

Palmer, C. J., Lawson, R. P., and Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*, *143*(5), 521–542. https://doi.org/10.1037/bul0000097

Pellicano, E., and Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Penzes, P., Cahill, M. E., Jones, K. A., VanLeeuwen, J.-E., and Woolfrey, K. M. (2011). Dendritic spine pathology in neuropsychiatric disorders. *Nature Neuroscience*, *14*(3), 285.

Phelan, K., and McDermid, H. E. (2011). The 22q13. 3 deletion syndrome (Phelan-McDermid syndrome). *Molecular Syndromology*, *2*(3–5), 186–201. https://doi.org/10.1159/000334260

Pina-Camacho, L., Parellada, M., and Kyriakopoulos, M. (2016). Autism spectrum disorder and schizophrenia: boundaries and uncertainties. *BJPscyh Advances*, *22*(5), 316–324. https://doi.org/10.1192/apt.bp.115.014720

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, *94*(5), 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018

Pitkänen, A., Pikkarainen, M., Nurminen, N., and Ylinen, A. (2000). Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postrhinal cortex in rat: a review. *Annals of the New York Academy of Sciences*, *911*(1), 369–391.

Piven, J., Elison, J. T., and Zylka, M. J. (2017). Toward a conceptual framework for early brain and behavior development in autism. *Molecular Psychiatry*, *22*(10), 1385–1394. https://doi.org/10.1038/mp.2017.131

Plaisted, K. C. (2000). Aspects of autism that theory of mind cannot explain. *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience (2nd Ed.).*, 222–250.

Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to Weak Central Coherence. *The Development of Autism: Perspectives from Theory and Research*, 149–169. https://doi.org/10.1017/CBO9781107415324.004

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of emotion* (pp. 3–33). https://doi.org/10.1016/b978-0-12-558701-3.50007-7

Polimanti, R., and Gelernter, J. (2017). Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genetics*, *13*(2), 1–14. https://doi.org/10.1371/journal.pgen.1006618

Quiroga, R. Q., and Kreiman, G. (2010). Measuring Sparseness in the Brain: Comment on Bowers (2009). *Psychological Review,* 117(1), 291-297.

Quiroga, R. Q., Kreiman, G., Koch, C., and Fried, I. (2008). Sparse but not "Grandmother-cell" coding in the medial temporal lobe. *Trends in Cognitive Sciences, 12*(3), 87–91. https://doi.org/10.1016/j.tics.2007.12.003

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

Pirowska, A., Wloch, T., Nowobilski, R., Plaszewski, M., Hocini, A., and Ménager, D. (2014). Phantom phenomena and body scheme after limb amputation: a literature review. *Neurologia i Neurochirurgia Polska*, 48(1), 52-59.

Ramachandran, V. S., and Seckel, E. (2015). Synesthetic colors induced by graphemes that have not been consciously perceived. *Neurocase, 21*(2), 216–219. https://doi.org/10.1080/13554794.2014.890728

Ramachandran, V. S., and Oberman, L. M. (2009). Broken Mirrors: A Theory of Autism. *Scientific American, 295*(5), 62–69. https://doi.org/10.1038/scientificamerican1106-62

Reddy, J. N. (2004). *An introduction to the finite element method* (Vol. 1221). New York, USA: McGraw-Hill.

Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3(11s), 1199.

Rizzolatti, G., and Fabbri-Destro, M. (2010). Mirror neurons: from discovery to autism. *Experimental Brain Research*, 200(3-4), 223-237.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131–141.

Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., and Baker, C. I. (2013). Slower Rate of Binocular Rivalry in Autism. *Journal of Neuroscience, 33*(43), 16983–16991. https://doi.org/10.1523/JNEUROSCI.0448-13.2013

Robertson, C. E., and Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn.2017.112

Robinson, E. B., Samocha, K. E., Kosmicki, J. A., McGrath, L., Neale, B. M., Perlis, R. H., and Daly, M. J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proceedings of the National Academy of Sciences, 111*(42), 15161–15165. https://doi.org/10.1073/pnas.1409204111

Rouw, R., and Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience, 10*(6), 792.

Rubenstein, J. L. R., and Merzenich, M. M. (2003). Model of autism : increased ratio of excitation / inhibition in key neural systems. *Brain, 2*(5), 255–267. https://doi.org/10.1046/j.1601-183X.2003.00037.x

Rutter, M. (1972). Childhood schizophrenia reconsidered. *Journal of Autism and Developmental Disorders, 2*(3), 315–337.

Sandin, S., Reichenberg, A., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., and Larsson, H. (2017). The heritability of autism spectrum disorder. *JAMA - Journal of the American Medical Association*, *318*(12), 1182–1184. https://doi.org/10.1001/jama.2017.12141

Sharma, A., Hoeffer, C. A., Takayasu, Y., Miyawaki, T., McBride, S. M., Klann, E., and Zukin, R. S. (2010). Dysregulation of mTOR Signaling in Fragile X Syndrome. *Journal of Neuroscience, 30*(2), 694–702. https://doi.org/10.1523/jneurosci.3696-09.2010

Sheng, M., and Kim, E. (2000). The Shank family of scaffold proteins. *Journal of Cell Science, 113 (Pt 1),* 1851–1856.

Smoller, J. W., Kendler, K., Craddock, N., Lee, P. H., Neale, B. M., Nurnberger, J. N., et al. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379. https://doi.org/10.1016/S0140-6736(12)62129-1

Sowden, S., Koehne, S., Catmur, C., Dziobek, I., and Bird, G. (2016). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research*, 9(2), 292-300.

Sokolowski, K., and Corbin, J. G. (2012). Wired for behaviors: from development to function of innate limbic system circuitry. *Frontiers in Molecular Neuroscience*, 5, 55.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., and Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nature Reviews Genetics*, *14*(7), 483–495. https://doi.org/10.1038/nrg3461

Song, L., Ma, H., Wu, M., Zhou, Z., and Fu, M. (2018). A Brief Survey of Dimension Reduction. *International Conference on Intelligent Science and Big Data Engineering* (pp. 189-200). Springer, Cham.

Soutschek, A., Burke, C. J., Raja Beharelle, A., Schreiber, R., Weber, S. C., Karipidis, I. I., et al. (2017). The dopaminergic reward system underpins gender differences in social preferences. *Nature Human Behaviour*, *1*(11), 819–827. https://doi.org/10.1038/s41562-017-0226-y

Spengler, S., Bird, G., and Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry*, 68(12), 1148-1155.

Spreen, O. B. A. L., and Benton, A. L. (1969). Embedded figures test. *Neuropsychological Laboratory, University of Victoria, British Columbia*, Canada.

Squire, L. R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Srivastava, S., and Sahin, M. (2017). Autism spectrum disorder and epileptic encephalopathy: common causes, many questions. *Journal of Neurodevelopmental Disorders*, *9*(1), 23. https://doi.org/10.1186/s11689-017-9202-0

Supekar, K., Kochalka, J., Schaer, M., Wakeman, H., Qin, S., Padmanabhan, A., and Menon, V. (2018). Deficits in mesolimbic reward pathway underlie social interaction impairments in children with autism. *Brain*, *141*(9), 2795–2805. https://doi.org/10.1093/brain/awy191

Sutton, R. S. (1984) Temporal credit assignment in reinforcement learning. PhD thesis, University of Massachusetts.

Südhof, T. C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. *Nature*, *455*(7215), 903–911. https://doi.org/10.1038/nature07456

Lőrincz, A., and Szirtes, G. (2009). Here and now: how time segments may become events in the hippocampus. *Neural Networks*, 22(5-6), 738-747.

Szita, I., and Lőrincz, A. (2009). Optimistic initialization and greediness lead to polynomial time learning in factored MDPs. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1001–1008. ACM.

Taube, J. S. (2007). The head direction signal: origins and sensory-motor integration. Ann. Rev. Neurosci. 30, 181–207. doi: 10.1146/annurev.neuro.29.051605.112854

Tick, B., Bolton, P., Happé, F., Rutter, M., and Rijsdijk, F. (2016). Heritability of autism spectrum disorders: A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *57*(5), 585–595. https://doi.org/10.1111/jcpp.12499

Toro, R., Konyukh, M., Delorme, R., Leblond, C., Chaste, P., Fauchereau, F., et al. (2010). Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends in Genetics*, *26*(8), 363–372. https://doi.org/10.1016/j.tig.2010.05.007

Troca-Marín, J. A., Alves-Sampaio, A., and Montesinos, M. L. (2012). Deregulated mTOR-mediated translation in intellectual disability. *Progress in Neurobiology, 96*(2), 268–282. https://doi.org/10.1016/j.pneurobio.2012.01.005

Tsakiris, M., Jiménez, A. T., and Costantini, M. (2011). Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations. *Proceedings of the Royal Society B: Biological Sciences, 278*(1717), 2470-2476.

van de Lagemaat, L. N., and Grant, S. G. N. (2010). Genome variation and complexity in the autism spectrum. *Neuron, 67*(1), 8–10. https://doi.org/10.1016/j.neuron.2010.06.026

Van Doorn, G., Hohwy, J., and Symmons, M. (2014). Can you tickle yourself if you swap bodies with someone else? *Consciousness and Cognition, 23*(1), 1–11. https://doi.org/10.1016/j.concog.2013.10.009

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning - ICML '08*, 1096–1103. https://doi.org/10.1145/1390156.1390294

Vorstman, J. A. S., Parr, J. R., Moreno-De-Luca, D., Anney, R. J. L., Nurnberger, J. I., and Hallmayer, J. F. (2017). Autism genetics: Opportunities and challenges for clinical translation. *Nature Reviews Genetics, 18*(6), 362–376. https://doi.org/10.1038/nrg.2017.4

Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412*(6842), 43.

Ward, J., Brown, P., Sherwood, J., and Simner, J. (2018). An autistic-like profile of attention and perception in synaesthesia. *Cortex, 107*, 121–130.

Warrier, V., Toro, R., Won, H., Leblond, C. S., Cliquet, F., Delorme, R., et al. (2019). Social and non-social autism symptoms and trait domains are genetically dissociable. Communications Biology, 2(1), 328.

Watson, B. O., and Buzsáki, G. (2015). Sleep, memory and brain rhythms. *Daedalus,* 144(1), 67-82.

Watson, T. L., and Robbins, R. A. (2014). The nature of holistic processing in face and object recognition: current opinions. *Frontiers in Psychology*, 5, 3.

Wilson, H. L., Wong, A. C. C., Shaw, S. R., Tse, W. Y., Stapleton, G. A., Phelan, M. C., et al. (2003). Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *Journal of Medical Genetics, 40*(8), 575–584.

Winter, S. S., and Taube, J. S. (2014). Head direction cells: from generation to integration, in S*pace, Time and Memory in the Hippocampal Formation*, (Berlin: Springer), 83–106.

Xiao, L., Priest, M. F., Nasenbeny, J., Lu, T., and Kozorovitskiy, Y. (2017). Biased Oxytocinergic Modulation of Midbrain Dopamine Systems. *Neuron, 95*(2), 368-384.e5. https://doi.org/10.1016/j.neuron.2017.06.003

Xu, C., Krabbe, S., Gründemann, J., Botta, P., Fadok, J. P., Osakada, F., et al. (2016). Distinct Hippocampal Pathways Mediate Dissociable Roles of Context in Memory Retrieval. *Cell, 167*(4), 961-972.e16. https://doi.org/10.1016/j.cell.2016.09.051

Yang, J. F., Lamont, E. V., and Pang, M. Y. (2005). Split-belt treadmill stepping in infants suggests autonomous pattern generators for the left and right leg in humans. *Journal of Neuroscience*, 25(29), 6869-6876. https://doi.org/10.1523/JNEUROSCI.1765-05.2005

Yasuhara, A. (2010). Correlation between EEG abnormalities and symptoms of autism spectrum disorder (ASD). *Brain and Development, 32*(10), 791–798. https://doi.org/10.1016/j.braindev.2010.08.010

Ye, K., Iossifov, I., Levy, D., Yamrom, B., Buja, A., Krieger, A. M., and Wigler, M. (2017). Measuring shared variants in cohorts of discordant siblings with applications to autism. *Proceedings of the National Academy of Sciences*, *114*(27), 7073–7076. https://doi.org/10.1073/pnas.1700439114

Yeung, K. S., Tso, W. W. Y., Ip, J. J. K., Mak, C. C. Y., Leung, G. K. C., Tsang, M. H. Y., Ying, D., Pei, S. L. C., Lee, S. L., Yang, W., and Chung, B. H. Y. (2017). Identification of mutations in the PI3K-AKT-mTOR signalling pathway in patients with macrocephaly and developmental delay and/or autism. *Molecular Autism*, *8*(1), 1–11. https://doi.org/10.1186/s13229-017-0182-4

Younes, L., Michor, P. W., Shah, J. M., and Mumford, D. B. (2008). A metric on shape space with explicit geodesics. *Rendiconti Lincei-Matematica e Applicazioni*, *19*(1), 25–57.

Zhao, X., Leotta, A., Kustanovich, V., Lajonchere, C., Geschwind, D. H., Law, K., et al. (2007). A unified genetic theory for sporadic and inherited autism. *Proceedings of the National Academy of Sciences*, *104*(31), 12831–12836. https://doi.org/10.1073/pnas.0705803104

Zhong, W., Li, Y., Huang, Y., Li, H., and Mo, L. (2018). Is the Lateralized Categorical Perception of Color a Situational Effect of Language on Color Perception? *Cognitive Science*, *42*(1), 350–364. https://doi.org/10.1111/cogs.12493