Mottron Laurent (Orcid ID: 0000-0001-5668-5422)

A radical change in our autism research strategy is needed: back to prototypes

Running title: **Autism: back to prototypes**

Laurent Mottron[a,b]

[a] Department of Psychiatry and Addictology, University of Montreal, Roger-Gaudry Pavillon, 2900 blvd Edouard-Montpetit, Montreal, QC, Canada, H3T 1J4

[b] CIUSSS-NIM Research Center, Rivière-des-Prairies, Mental Health Hospital, 7070 blvd Perras, Montreal, QC, CANADA, H1E 1A4

Corresponding Author:

Laurent Mottron

Tel: 1-514-972-4841

E-mail: laurent.mottron@gmail.com

Laurent Mottron, Riviere-des-Prairies Mental Health Hospital,

7070 blvd Perras, Montreal, Quebec, H1E 1A4.

ROLE OF FUNDING SOURCE

Funding for this study was provided by the University of Montreal M&R Gosselin Research Chair on Cognitive Neuroscience of Autism. The Chair had no role in the study design, collection, analysis or interpretation of the data, writing of the manuscript, or the decision to submit the paper for publication.

CONTRIBUTOR

Laurent Mottron conducted literature searches, provided summaries of previous research studies, and wrote the final manuscript.

CONFLICT OF INTEREST

Laurent Mottron declares that he has no conflict of interest.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/aur.2494

LAY SUMMARY

Scientific research into the causes of autism and its mechanisms is carried out on large cohorts of people who are less and less different from the general population. This historical trend may explain the poor harvest of results obtained. Services and intervention are provided according to a diagnosis that now encompasses extremely different individuals. Lastly, we accept as a biological reality the constant increase over the years in the proportion of autistic people among the general population. These drifts are made possible by the attribution of a diagnosis of autism to people who meet vague criteria, rather than to people who experienced clinicians recognize as autistic. We propose to change our research strategy by focusing on the study of the latter, fewer in number, but more representative of the "prototype" of autism. To do this, it is necessary to clearly distinguish the population on which the research is carried out from that to which we provide support. People must receive services according to their needs, and not according to the clarity of their diagnosis.

ABSTRACT

The evolution of autism diagnosis, from its discovery to its current delineation using standardized instruments, has been paralleled by a steady increase in its prevalence and heterogeneity. In clinical settings, the diagnosis of autism is now too vague to specify the type of support required by the concerned individuals. In research, the inclusion of individuals categorically defined by over-inclusive, polythetic criteria in autism cohorts results in a population whose heterogeneity runs contrary to the advancement of scientific progress. The search for models common to individuals sharing only a trivial resemblance produces a large-scale type-2 error (not finding differences between minor phenotypic variants and the dominant population) rather than resulting in the detection of mechanistic differences to explain phenotypic differences between them. However, the dimensional approach of autism proposed to cure the disease of its categorical diagnosis is plagued by the arbitrariness of the choice of relevant dimensions under study. Here, we argue that an emphasis on the reliability rather than specificity of diagnostic criteria and the misuse of diagnostic instruments, which ignore the recognition of a prototype, place autism at a steadily higher level in the hierarchical taxonomy of neurodevelopmental variants. We propose a remedy in which centering research on cohorts in which individuals are selected based on their expert-judged prototypicality may represent a theoretical and practical resolution of certain pervasive issues pertaining to autism diagnostic thresholds. Reversing the

current research strategy by giving more weight to specificity than reliability should increase our ability to discover the mechanisms of autism.

KEYWORDS: reliability, type 2 error, prototype, diagnostic, polythetic criteria.

We should not feel triumphant about the advances in cognitive neuroscience, genetics, or brain imaging or our general understanding of the etiological aspects of non-syndromic autism made over the past 30 years. Hence, we are pessimistic about the prospect of big future breakthroughs in our mechanistic understanding of autism. The reason for such pessimism is that recent research practices and methodological norms have overwhelmingly favored the production of type 2-like errors, thus not detecting mechanisms that account for the nature and existence of autism. Recent meta-analytical studies indicate that case-control effect sizes have decreased by up to 80% for neurocognitive constructs (emotional recognition, planning, capacity of cognitive perspective taking, brain size, and EEG characteristics) that distinguish autistic from non-autistic people (Rodgaard, Jensen, Vergnes, Soulieres, & Mottron, 2019). The gradual 30-fold increase in the prevalence of people diagnosed as autistics over the last 50 years coincides with the inclusion of individuals who are increasingly distant from the initial description (Fombonne, 2018; Hollin, 2017), resulting in increasing heterogeneity. The number of signs required to provide an autism diagnosis decreased by a factor of two between 2004-2005 and 2014 for children diagnosed at school age in Sweden (Arvidsson, Gillberg, Lichtenstein, & Lundström, 2018).

The evolution in the demarcation of autism and the detection of the difference between autistic and non-autistic individuals has been accompanied by minimal replicability of structural and functional results in brain imaging. In genetics, the most important results are those that have ruled out an important causal role of entire classes of genetic abnormalities (such as deletions: Douard et al. (2020). Concerning interventions, the major findings have been the negative results that show the minimal or dubious effectiveness of intervention techniques (Brignell et al., 2018; Sandbank et al., 2020).

Some researchers have suggested breaking down the autism spectrum into subgroups to treat this ailment. However, meta-analyses of studies attempting to create subgroups for the current autism spectrum report that the number of possible clusters may be impractically large, with most of doubtful clinical value (Wolfers et al., 2019). Concluding that this demonstrates the validity of the spectrum category (Fombonne, 2020) may miss the point. The current

dilemma may instead be explained by the current definition of autism spectrum not allowing the detection of subgroups because it gathers unrelated and dissimilar sets of individuals.

Proponents of a dimensional position see such drift as progress. The sharing of diagnostic signs with other multiple psychiatric and neurodevelopmental conditions and the existence of common predisposing factors between autism and these same conditions could suggest that such a categorical distinction has become obsolete (Constantino & Charman, 2016). However, although categories are plagued by the problem of boundary, dimensions suffer from a problem of choice. The use of dimensional measures to treat the reification of disease substitutes the grouping of individuals into a category deemed to be arbitrary, with the classification of individuals according to the measure of a dimension, of which the choice is even more arbitrary. Another intrinsic limit of the dimensional approach is the uncontrollable increase in the number of dimensions when the complexity of objects increases, or the "curse of dimensionality" (Feczko et al., 2019). Such assimilation confuses the possibility of measuring the same variable, such as reciprocal socialization (Constantino et al., 2003) or empathizing/systemizing pairs (Baron-Cohen, 2009) in all individuals of a group, and its explanatory value in a mechanistic model.

There may be several causes at the center of the current situation: the standardization of inclusion strategies for individuals identified as autistic in research, the blind application of inappropriate methodological rules, such as requiring a large sample size and the search for representativeness, an inappropriate sign/specifier distinction, and a misuse of the pleiotropism analogy to autism without identified variants. They will be discussed in this order.

Methodological dogmas, premature assumptions, and case ascertainment strategies that contribute to the trivialisation of autism

Reliability/standardization: The dogma for the diagnosis of autism is the use of validated and standardized instruments that unify the operationalization of DSM criteria and reduce the discrepancy between individual judgments. We suspect such standardisation of diagnostic procedures to be largely responsible for the plateauing of autism research, by adding artifactual or criteria- or instrument-based heterogeneity to the natural variability of autistic presentation

due to sex, age and outcome. The diagnosis of autism is obtained using these instruments when reaching a threshold summary score by adding individual item scores (Randall et al., 2018). Their cut-off threshold scores are determined by a specificity-sensitivity trade-off, expert agreement long ago being their reference. Multiple warnings, especially by C. Lord, that they should not be used alone and without a clinical judgment have been essentially abolished by their commercial presentation as diagnostic instruments. However, we now know that such instruments are over-inclusive (Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011), influenced by non-specific dimensions (Fombonne et al., 2020; Havdahl et al., 2016), and vulnerable to large-scale temporal evolution (Arvidsson et al., 2018). Despite such warnings, most research papers use them as an entry point without further refinement. Autism in the clinical and research world of today is what is measured by the ADI-R and ADOS-G and reliability is confused with truth.

The issue with standardized instruments may be intrinsic to the use of summary scores of polythetic criteria. Summary scores privilege the grouping of exemplars that share certain features –trivial when quantitatively measured- over the intersection of maximally resembling exemplars. Moreover, signs in standardized instruments are independent to avoid a ''halo effect'', that is the bias to detect one sign when another related one is present – the negative counterpart of expertise. Therefore, their grouping into ''metasigns'', subsets of qualitatively specified signs that strengthen the clinical recognition of the diagnosis when present together, is lost in the operation. Furthermore, signs in polythetic systems are not differentially weighted: their contribution to the varying distance from the prototype is replaced by a global quantitative pass or fail. Overall, polythetic criteria and their ascension in a hierarchical classification increase reliability, but at the risk of turning it into triviality: if judges are divided to decide whether a bumpy circle is indeed a circle, they will all agree that both are shapes. The abstract nature of certain DSM 5 criteria of autism (e.g., A3: deficits in developing and maintaining relationships) is a dramatic example of reliability turned into triviality.

Sample size. A conviction shared by the scientific community in autism is that the first research on small samples biased the results in favor of their initial hypotheses, whereas studies on a large N, with high standards, brought the previously found results into a more just

light. This belief is consistent with the belief that meta-analyses provide us with a safer message than individual studies. However, there are undeniable examples (e.g. in intervention: Pickles et al., 2016) in which a single study is better than a thousand studies with lower standards (Dawson & Fletcher-Watson, 2020). Moreover, in the current state of the definition of the autism spectrum, the primacy attributed to the size of the sample over the resemblance of the individuals who compose it creates a level of noise that increases dramatically with the size of the sample. The avoidance of the type-1 risk associated with small samples must be balanced against the type-2 risk associated with large, heterogeneous samples. The phenomenon of Simpson's paradox (Pearl & Mackenzie, 2018) similarly describes the excessive weight given to outliers in a small sample and that of a diverging subgroup within a large sample. If the reliance on a large N is at the cost of an uncontrolled increase in heterogeneity, the gain obtained by increasing the N will be more than offset by the loss of information resulting from the noise of heterogeneity.

Representativeness. Ensuring the representativeness of the sample tested for the population under study will always prompt us to favor probability sampling over convenience sampling. However, random sampling within large cohorts is obtained at the cost of a constant rise in the hierarchical taxonomy of neurodevelopmental conditions, with which autism then becomes confused: any neurodevelopmental or adult psychiatric condition is now suspected to have autistic traits. Probability sampling only makes sense for a population for which the identification is unquestionable and can be taken as a starting point for research. Probability sampling provides false security while maximizing the biases or uncertainty introduced by the state of the description of a condition at a particular moment in time. The findings of autism research built on probabilistic sampling are only as good as the inclusion criteria used to select the population. Conversely, convenience sampling may target a specific question, limiting its representativeness to the question and the population under study. In this situation, I would be more confident in convenience sampling than probability sampling based on a large N that is poorly representative of the subgroup for which my question was initially asked.

The generalisability of a scientific result obtained with sample A to sample B is determined by the level of similarity between the individuals composing samples A and B. We

should not always aim for representativeness for the entire ''spectrum'': the true alternative is not between a chosen and a random sample, but between a sample chosen by a few experts for a particular purpose vs. that chosen by consensual, universal, and all-purpose criteria, which have, after 30 years, produced minimal decisive knowledge. Similarly, the dogma that a meta-analysis is by definition more informative than a single study needs to be challenged. A meta-analysis is worth what the studies that compose it are worth, but it is less generalizable than an individual study involving a sample that complies with the question asked.

Justifying ''spectrum'' by pleiotropism. The variability of the presentation of monozygotic twins concordant for autism teaches us that a strictly identical genetic predisposition can produce different pictures, which could be taken as an argument against the concept of protoypicality. It defines a certain type of variability, "from genes to behavior", although the nature of the genetic predisposition remains unknown. However, we see several issues in the premature assimilation between autism with and without identified variants, which limits the use of ''pleiotropism'' as a counterexample of prototypicality in the second situation. The concordance between twins increases with accepted prevalence, and hence diminishes when the "stringency" of the diagnostic criteria increases (Tick et al. 2016). The very concept of concordance looses its relevance under a certain density of ''autistic'' traits. It is precisely the legitimization of heterogeneity by pleitropism with which we disagree: the current vagueness of diagnostic criteria and modes of inclusion in cohorts to encompass more than the variability observed in a situation validated by diagnosis concordance *and* by phenotypical recognition.

To advance the detection of the genetic mechanisms at play in autism without identified variants, we must leave open the nature of the relationship between the autism phenotype at the center of the category and at its periphery (Fish, 2017). Pleiotropy is a legitimate research topic to be integrated into ''gene-to-behavior'' mechanistic models when the genetic variant is identified. However, the multiplicity of observed phenotypes associated with a family of neighboring mutations (e.g. 22q11) in a direct causality scheme has a similar – but considerably amplified - correspondent in the multiplicity of causality models consistent with a ''spectrum'' phenotype in a ''reverse causality'' scheme (see Figure 1a). Any increase in

the variability of this phenotype has a multiplicative effect on its possible causes, as any enlargement of the autism definition from prototype to polythetic criteria has a multiplicative effect on case ascertainment prevalence and heterogeneity (Figure 1b). Acknowledged pleitropism cannot justify the inclusion of heterogeneous individuals when we start from an ill-defined phenotype to identify un unknown genetic alteration (from behavior to gene).

The ''extreme of a distribution'' dogma: Another belief accepted as an advance of the last decade of research is that clinical autism is the extreme of a continuous distribution of autistic traits in the general population (Happe & Frith, 2020). However, the notion of autistic traits does not provide information about autism until the relationship between the individuals in whom these traits are measured is independently validated. Autistic traits are not autism, but rather social or cognitive features for which the relationship with recognizable autism can be interpreted as such only in limited situations in which they may indicate genetic predisposition to autism (Mottron & Bzdok, 2020). A genetic predisposition to autism is not autism, in the sense that each of these two levels constitute the stabilization of processes that each have their own logic, even if they are genetically related. Merging the two concepts confuses the chain of causality with the stability of the effects. Promoting a dimensional approach to autism is an undue conclusion to a verified fact, the existence of a broader phenotype. It has resulted in the detection of trivial similarity in an indefinite number of contexts and conditions. Accordingly, the selection of dimensions chosen as ''traits'' among the various manifestations of prototypical autism is arbitrary and extremely vague. Their anomalies are inherently non-specific (what psychiatric condition does not alter ''social reciprocity?'').

We therefore suggest a conceptual and operational separation between autism and autistic traits. This would lead to the use of the notion of 'autistic trait' only for people related to prototypical autism, thereby to the broader autistic phenotype. Even in this limited context, the reduction of the phenotype by promoting a single quantitative variable creates, circularly, an artifactual continuous distribution. This should be replaced by a systematic search for qualitative signs evident in relatives of an autistic prototypical proband. These are not necessarily ''autistic'', as shown by the presence of a simple and reversible language delay in siblings of autistic children (Marrus et al., 2018).

The distinction between signs and specifiers: Following the choice provided by the DSM 5, clinical specifiers can take any value without changing the belonging of an individual to the "autism spectrum" category. It is certainly possible to observe variations in measured intelligence, co-morbidities, or language, particularly during development (Georgiades, Bishop, & Frazier, 2017) of an initially prototypical phenotype. However, the membership of these different phenotypic variations in the autism category must be validated by an initial prototypical presentation. For example, in the case of the specifier "language", the objective heterogeneity of the outcomes of nonverbal autism during the preschool period is evident, but there is little chance of finding the mechanisms that prevail in autistic language development if we collectively study people who have different developmental trajectories *ab initio*.

Differential diagnosis has disappeared from the DSM criteria, replaced by the specifier ''comorbidity'', and was never incorporated in their operationalization in standardized tools. This decision could appear to be supported by the trivial discovery of autistic traits in an indefinite variety of psychiatric and neurodevelopmental conditions. We contend however that this DSM decision *produces* such a result and the thresholds of standardized instruments *rubberstamp* it. This dissolves autism into a now uncontrollable morass of heterogeneity. Conversely, the autism prototype is characterized by associations between values of the clinical specifiers - which are not independent of each other (e.g., Manelis et al. (2020). Because of the dilution of autistic signs in quantitative measures of abstract and nonspecific dimensions, one is now unable (even discouraged) to distinguish between comorbidity and a phenocopy - for example between social atypia linked to ADHD and to autism. In this situation, an identified co-morbidity in autistic people should not be considered a priori as informative for autism. This applies particularly to autism with neurogenetic conditions, which should not be considered as reciprocally informative with autism without these conditions, as being common members of an "autistic spectrum''. This distinction, although clear 20 years ago, has been gradually blurred. It took 20 years to (re) discover that autism with and without non-verbal intellectual disability differ (e.g., in heritability: Xie et al. (2020) to the point that they may be minimally reciprocally informative.

Contribution of prototype theory

Reaching a cut-off is "grouping without resemblance'', the opposite of the graded *familial resemblance* that characterizes prototypes (Wittgenstein 1953, Rosch 1978). There is an epistemic conflict between matching-to-prototype recognition, which is intrinsically graded, and a pass-or-fail diagnostic threshold (you are or are not autistic), which abolishes this gradation within a category. While prototype is based on family resemblance, the latter approach copies the necessary-sufficient framework that has proven to be appropriate only for mathematical fields, to work poorly in biology, and to have no psychological validity. It corresponds to a formal model of categorisation that fits neither with the way autism was discovered nor the psychological laws governing the use of concrete or abstract semantic entities when identifying a cluster of signs.

The application of prototype theory to an autism diagnosis grades the similarity of an individual to the subjective prototype of a limited number of experts who have long been exposed to an enriched population with suspected autism. We suggest replacing the reliance on increasing the N of studies incorporating heterogeneous individuals with increasing the N to whom the experts supervising recruitment have been exposed. What we lose in statistical power will be offset by a better signal-to-noise ratio, resulting from studying more resembling individuals.

Beyond the issue of gradation of familial resemblance, the notion of a prototype is associated with that of a basic level in a semantic hierarchy, in which the category maximises the information conveyed by correlated features. It coincides with the most frequently or precociously encountered set of features that discriminates one category from another at the same level. The mental image they evoke reflects the entire category without further analysis. The validity, the probability that a feature x predicts a category y and is not associated with another category, is maximized at the basic level. This notion of a basic level can be fruitfully applied to psychiatric categories, which are organized hierarchically (Flanagan, Keeley, & Blashfield, 2012). Superordinate categories (e.g., autism spectrum) share most of their attributes with contrasting categories (e.g., other neurodevelopmental conditions), thereby favoring diagnostic substitution.

The notion of basic level may address the problem of comorbidity, so frequent in autism (Hossain et al., 2020; Lai et al., 2019; Romero et al., 2016) that some have raised the question of the distinction between autism and other clinical entities. The current reported psychiatric comorbidity of autism may directly depend on the prototypicality of individuals on which it is calculated. A heterogeneous category, such as ''autism spectrum'', may be one that is simply too general, abstract, or superordinate and in need of decomposition into more homogeneous and informative categories by descending one level of the hierarchy. This was obvious in the overlap of the previous DSM IV category of PDDNOS, with an indefinite variety of neurodevelopmental and psychiatric conditions (de Bruin, Ferdinand, Meester, de Nijs, & Verheij, 2007).

Conclusion: priority should be given to prototypes in research populations

We hypothesize that people judged to be the most prototypical *in presentia* by different judges will be more similar to each other than those meeting an identical threshold of a standardized instrument. Such a population would be arbitrarily truncated to the power sufficient to obtain a difference in autism compartment under study, so as to keep only the most prototypical individuals of each major subgroup for a determined N. It is not a question of keeping the most severe individuals, because severity, as currently defined, confuses adaptation and intelligence.

We therefore propose the following steps for the creation of a research cohort combining the advantages of standardized categorical type diagnosis and gradation in prototypicality. a) Sample a population that exceeds the sum score of a standardized threshold. b) Decompose the population into compartments with homogeneous values for the DSM 5 specifiers (e.g., *comorbidity*: with vs without CNV or neuro-genetic conditions; *language*: with vs without initial language delay; *intelligence*: with vs without non-verbal intellectual disability) to which will be added *age* (preschool vs. school and adult age) and *sex*. c) Classify in situ individuals who make up these compartments by decreasing prototypicality. This ranking is obtained by averaging the score of each participant according to two experts based on the following elements: level of similarity to his personal autism category, speed of clinical

identification, exemplarity for academic teaching. d) Determine an N sufficient for the desired power and truncate the compartments to these Ns. e) Finally, compare the case-control differences obtained in each of these compartments to test their generalizability.

We are at a time in the advancement of science in which we ignore the delineation of autism, but in which its consensual definition constrains our ability to design mechanistic models. The autism spectrum as currently defined by the DSM and operationalized by standardized tools should not be the starting point for scientific research in neuroscience. The autism spectrum is a convention that changes over time and belongs more to the history of science than neurobiology, while limiting the discovery of this latter discipline. The discovery record of the dimensional alternative, regularly offered as an alternative to categorical diagnostic decision-making, is even bleaker. The choice of considering dimensions to be relevant is even more arbitrary than that of the limits of categories. Although it means formalizing an arbitrary limit to the inclusion of a person in a research cohort, focusing research on prototypical individuals must result in populations that favor the production of knowledge – which is not currently the case.

References

Arvidsson, O., Gillberg, C., Lichtenstein, P., & Lundström, S. (2018). Secular changes in the symptom level of clinically diagnosed autism. *J Child Psychol Psychiatry, 59*(7), 744-751. doi:10.1111/jcpp.12864

Baron-Cohen, S. (2009). Autism: the empathizing-systemizing (E-S) theory. *Ann N Y Acad Sci, 1156*, 68-80.

Brignell, A., Chenausky, K. V., Song, H., Zhu, J., Suo, C., & Morgan, A. T. (2018). Communication interventions for autism spectrum disorder in minimally verbal children. *Cochrane Database Syst Rev, 11*, CD012324. doi:10.1002/14651858.CD012324.pub2

Constantino, J. N., & Charman, T. (2016). Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. *Lancet Neurol, 15*(3), 279-291. doi:10.1016/S1474-4422(15)00151-9

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., . . . Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J Autism Dev Disord, 33*(4), 427-433. Retrieved from file://W:%5Cautisme%20et%20langage%5Cbiblio%5Cconstantino2003%20SRS.pdf

Dawson, M., & Fletcher-Watson, S. (2020). Commentary: What conflicts of interest tell us about autism intervention research-a commentary on Bottema-Beutel et al. (2020). *J Child Psychol Psychiatry*. doi:10.1111/jcpp.13315

de Bruin, E. I., Ferdinand, R. F., Meester, S., de Nijs, P. F., & Verheij, F. (2007). High rates of psychiatric co-morbidity in PDD-NOS. *J Autism Dev Disord, 37*(5), 877-886. doi:10.1007/s10803-006-0215-x

Douard, E., Zeribi, A., Schramm, C., Tamer, P., Loum, M. A., Nowak, S., . . . Jacquemont, S. (2020). Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. *Am J Psychiatry*, appiajp202019080834. doi:10.1176/appi.ajp.2020.19080834

Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A. M., Nigg, J. T., & Fair, D. A. (2019). The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends Cogn Sci, 23*(7), 584-601. doi:10.1016/j.tics.2019.03.009

Fisch GS. Whither the genotype-phenotype relationship? An historical and methodological appraisal. Am J Med Genet C Semin Med Genet. 2017 Sep;175(3):343-353. doi: 10.1002/ajmg.c.31571. Epub 2017 Aug 21. PMID: 28834162.

Flanagan, E., Keeley, J., & Blashfield, R. (2012). Do clinicians conceptualize DSM-IV disorders hierarchically? *J Clin Psychol, 68*(6), 620-630. doi:10.1002/jclp.21840

Fombonne, E. (2018). Editorial: The rising prevalence of autism. *J Child Psychol Psychiatry, 59*(7), 717-720. doi:10.1111/jcpp.12941

Fombonne, E. (2020). Camouflage and autism. *J Child Psychol Psychiatry, 61*(7), 735-738. doi:10.1111/jcpp.13296

Fombonne, E., Morotti, H., Mastel, S., Keller, K., Barnard, R. A., Hall, T., & O'Roak, B. J. (2020). Autism questionnaire scores do not only rise because of autism. *Dev Med Child Neurol*. doi:10.1111/dmcn.14725

Georgiades, S., Bishop, S. L., & Frazier, T. (2017). Editorial Perspective: Longitudinal research in autism - introducing the concept of 'chronogeneity'. *J Child Psychol Psychiatry, 58*(5), 634-636. doi:10.1111/jcpp.12690

Happe, F., & Frith, U. (2020). Annual Research Review: Looking back to look forward - changes in the concept of autism and implications for future research. *J Child Psychol Psychiatry, 61*(3), 218-232. doi:10.1111/jcpp.13176

Havdahl, K. A., Hus Bal, V., Huerta, M., Pickles, A., Oyen, A. S., Stoltenberg, C., . . . Bishop, S. L. (2016). Multidimensional Influences on Autism Symptom Measures: Implications for Use in Etiological Research. *J Am Acad Child Adolesc Psychiatry, 55*(12), 1054-1063 e1053. doi:10.1016/j.jaac.2016.09.490

Hollin, G. (2017). Autistic Heterogeneity: Linking Uncertainties and Indeterminacies. *Sci Cult (Lond), 26*(2), 209-231. doi:10.1080/09505431.2016.1238886

Hossain, M. M., Khan, N., Sultana, A., Ma, P., McKyer, E. L. J., Ahmed, H. U., & Purohit, N. (2020). Prevalence of comorbid psychiatric disorders among people with autism spectrum disorder: An umbrella review of systematic reviews and meta-analyses. *Psychiatry Res, 287*, 112922. doi:10.1016/j.psychres.2020.112922

Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., . . . Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *Lancet Psychiatry, 6*(10), 819-829. doi:10.1016/S2215-0366(19)30289-5

Manelis, L., Meiri, G., Ilan, M., Flusser, H., Michaelovski, A., Faroy, M., . . . Menashe, I. (2020). Language regression is associated with faster early motor development in children with autism spectrum disorder. *Autism Res, 13*(1), 145-156. doi:10.1002/aur.2197

Marrus, N., Hall, L. P., Paterson, S. J., Elison, J. T., Wolff, J. J., Swanson, M. R., . . . Network, I. (2018). Language delay aggregates in toddler siblings of children with autism spectrum disorder. *J Neurodev Disord, 10*(1), 29. doi:10.1186/s11689-018-9247-8

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism, 15*(2), 143-162. doi:10.1177/1362361310379241

Mottron, L., & Bzdok, D. (2020). Autism spectrum heterogeneity: fact or artifact? *Mol Psychiatry*. doi:10.1038/s41380-020-0748-y

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. New York: Basic Books.

Pickles, A., Le Couteur, A., Leadbitter, K., Salomone, E., Cole-Fletcher, R., Tobin, H., . . . Green, J. (2016). Parent-mediated social communication therapy for young children with autism (PACT): long-term follow-up of a randomised controlled trial. *Lancet, 388*(10059), 2501-2509. doi:10.1016/s0140-6736(16)31229-6

Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J., Hooft, L., Livingstone, N., . . . Williams, K. (2018). Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database Syst Rev, 7*, CD009044. doi:10.1002/14651858.CD009044.pub2

Rodgaard, E. M., Jensen, K., Vergnes, J. N., Soulieres, I., & Mottron, L. (2019). Temporal Changes in Effect Sizes of Studies Comparing Individuals With and Without Autism: A Meta-analysis. *JAMA Psychiatry, 76*(11), 1124-1132. doi:10.1001/jamapsychiatry.2019.1956

Romero, M., Aguilar, J. M., Del-Rey-Mejias, A., Mayoral, F., Rapado, M., Pecina, M., . . . Lara, J. P. (2016). Psychiatric comorbidities in autism spectrum disorder: A comparative study between DSM-IV-TR and DSM-5 diagnosis. *Int J Clin Health Psychol, 16*(3), 266-275. doi:10.1016/j.ijchp.2016.03.001

Rosch, E. (1978). Principes of Categorization. In E. Rosch, and Lloyd, B.B. (Ed.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum.

Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J. I., . . . Woynaroski, T. G. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychol Bull, 146*(1), 1-29. doi:10.1037/bul0000215

Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry. 2016 May;57(5):585-95. doi: 10.1111/jcpp.12499. Epub 2015 Dec 27. PMID: 26709141; PMCID: PMC4996332.

Wolfers, T., Floris, D. L., Dinga, R., van Rooij, D., Isakoglou, C., Kia, S. M., . . . Beckmann, C. F. (2019). From pattern classification to stratification: towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neurosci Biobehav Rev, 104*, 240-254. doi:10.1016/j.neubiorev.2019.07.010

Xie, S., Karlsson, H., Dalman, C., Widman, L., Rai, D., Gardner, R. M., . . . Lee, B. K. (2020). The Familial Risk of Autism Spectrum Disorder with and without Intellectual Disability. *Autism Res*. doi:10.1002/aur.2417

Figure legend

Figure 1a & b: Variation in heterogeneity as a function of directionality in diagnosis and research

1a. Directional heterogeneity in research. In a causal direction from genes to behavior, pleiotropy describes the increase in variability x, from that of the known genetic cause (x) to that of the corresponding phenotypic deviation (nx). In a genetic investigation or imaging by reverse causality, from behavior or cognition, this increase in variability is disproportionately increased ($nx \rightarrow nx^n$), resulting in non-replicability of the results.

1b. Directional heterogeneity in diagnosis. The historical evolution from initial recognition of the prototype to identification by criteria results in an increase in the variability of the phenotype ($x \rightarrow nx$). In the reverse direction (from diagnostic criteria to case-ascertainment), the variability increases disproportionately ($nx \rightarrow nx^n$), producing an epidemic increase in prevalence.