Modeling aberrant volatility estimates in Autism Spectrum Disorder

Hauke Niehaus[1], Sanna Stroth[2], Inge Kamp-Becker [2], Dominik Endres[1]
[1]Department of Psychology, Philipps-University Marburg
[2]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Medical Clinic, Philipps-University Marburg
{niehaush, stroth, endresd}@staff.uni-marburg.de, {kampbeck}@med.uni-marburg.de

Abstract

Computational cognitive theories of Autism Spectrum Disorder have received renewed attention in recent years. Consistent with the predictive processing framework, ASD has been re-conceptualized as a disorder of aberrant prediction and learning-rate estimation involving multiple levels of a putative cognitive computational hierarchy. Specifically, behavioral symptoms of individuals with ASD might manifest due to an aberrant overestimation of the volatility of environmental contingencies (i.e. tendency of change in cue-outcome probabilities) which in turn might induce a dysfunctional setting of learning rates. In this work, we attempted to conceptually replicate computational modeling analyses of an impactful study of the recent ASD modeling literature in an independent sample of subjects. We were not able to replicate some prior reported effects likely due to differences in model architecture and cognitive task setup. We found statistical trends in similar directions.

Keywords: autism spectrum disorder; predictive processing; cognitive modeling; Wisconsin card sorting test; volatility

Introduction

Autism-Spectrum-Disorder (ASD) is currently classified as a Pervasive Neurodevelopmental Disorder with a point-prevalence of 1% in the general population (Happé & Frith, 2020; Lord & Bishop, 2015) with significant genetic associations (de La Torre-Ubieta, Won, Stein, & Geschwind, 2016; Waye & Cheng, 2018). The diagnosis of ASD requires the presence of symptoms from two main symptom clusters. (A) persistent and context-invariant deficits in social communication and interaction including socio-emotional reciprocity, nonverbal behavior, and relationship formation. (B) repetitive patterns of behavior, interests, or activities including motor movements, speech or use of objects, a strong insistence on routines and sameness, fixated and singular interests, and peculiar differences in sensory processing (*Diagnostic and statistical manual of mental disorders: DSM-5*, 2013).

Until recently, pathomechanistic and etiological theories of ASD were focused on domain-specific theories attempting to explain the symptom clusters separately like e.g. the weak-central coherence theory (Happé, 2005) and the enhanced perceptual functioning theory (Mottron, Dawson, Soulières, Hubert, & Burack, 2006) for perceptual peculiarities or e.g. the extreme male brain theories (Baron-Cohen, 2002; Greenberg, Warrier, Allison, & Baron-Cohen, 2018) for behavioral atypicalities. A coherent framework linking the symptoms clusters was largely missing (Haker, Schneebeli, & Stephan, 2016).

More recently, attempts have been made to unify all behavioral and perceptual atypicalities of ASD using the Bayesian Brain hypothesis (BB) (Lawson, Rees, & Friston, 2014; Pellicano & Burr, 2012; van de Cruys et al., 2014). In essence, the BB postulates that the human brain performs Bayesian inference through a generative model of the world and represents information via probability density functions or approximations thereof (Knill & Pouget, 2004). More specifically, Pellicano and Burr (2012) hypothesized on a computational level that people with ASD rely in their inferences about the world less on their learned prior information ('hypo-prior') than normally developing individuals. Formally, they proposed that the construction or updating of Bayesian priors (i.e. the contextually sensitive expectations of sensory inputs) might be aberrant in ASD leading to attenuated and overly imprecise prior expectations. This in turn manifests in a maladaptive perceptual style largely unbiased by the priors and closely oriented toward the sensory input. As the prior generally serves an important function in reducing uncertainty in Bayesian formulations of perception, Pellicano and Burr (2012) hypothesized that this aberrant construction of priors in ASD might lead to context-invariant and broad expectations necessarily leading to generally increased differences between prior expected and observed sensory input. Many ASD-associated perceptual and behavioral phenomena may be explained by this aberrant computational process (Haker). Lawson et al. (2014) adapted this computational hypothesis of Pellicano and Burr (2012) into the neurobiologically constrained predictive coding framework (also Predictive Processing, from here on PP) and extended it to the temporal and hierarchical dimensions. PP is a process theory that describes how computational Bayesian inference might be realized in the human brain (Clark, 2013). A central role falls to top-town predictions that attempts to explain away neuronal activity at lower levels of the neurocomputational hierarchy. A mismatch between top-down predictions and bottom-up activity leads to the generation of a weighted prediction error (PE) that is passed up the hierarchy to improve predictions. Importantly, PP departs from the representational framework of cortical computation by proposing that only PEs are passed upwards the cortical hierarchy (Keller & Mrsic-Flogel, 2018). Moreover, the (contextual) weighting of PEs is important in this scheme and, if estimated correctly, functions by discerning between irreducible noise in the environment and relevant contingencies. Finally, more shallow/bottom-up layers

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

of the proposed neurocomputational hierarchy are assumed to encode domain-specific PEs regarding temporally and spatially 'smaller' environmental aspects whereas higher levels encode domain-general PEs over temporally and spatially abstract environmental aspects (Clark, 2013; Perrykkad & Hohwy, 2020).

Specifically for ASD, Lawson et al. (2014) hypothesized that the aberrant Bayesian inference as described by Pellicano and Burr (2012) manifests in a relatively increased subjective precision (inverse variance) of sensory evidence relative to top-down prior beliefs, especially in moments of high ambiguity or uncertainty. This prevents the adequate weighting of PEs along the cortical hierarchy rendering them generally less informative in updating the generative model. This effect might be more pronounced the more domain-general the PEs become. This failure to accurately inform the generative model with only relevant prediction errors, especially on higher levels, might additionally prevent the formation of adequately abstract (or 'deep') generative models over time. As the accurate PE-weight (and precision) estimation requires contextual knowledge of the environment, ASD has consequently been described as a meta-cognitive disorder (Friston, Lawson, & Frith, 2013). In summary, many characteristic ASD symptoms might be compensatory behavioral mechanisms to reduce the aberrant PE weighting. For examples see e.g. (Haker et al., 2016).

After this introduction, we will summarize the methods employed in this study. Then, the main findings of Lawson, Mathys, and Rees (2017) will be summarized and our hypotheses based on those original results will be presented. Then, findings from our study will be reported and replication success will be critically evaluated. Finally, we will give an outlook for further research.

Methods

Subjects and demographic data

Subjects in our sample were screened for Autism between the years 2001 and 2009 as part of a research project on high-functioning autism and Asperger syndrome at the University Hospital Marburg. All subjects were classified as ASD or non-ASD cases based on best-estimate clinical (BEC) diagnosis according to ICD-10, comprising a comprehensive clinical investigation with physical examination, medical history-taking, assessment of intellectual ability with the HAWIK-3 (Tewes, Rossmann, & Schallberger, 1999), ADOS, ADI-R and differential diagnostic examination. Additionally, subjects completed several neuropsychological tests including the Wisconsin Card Sorting test. Informed consent was obtained either from the subjects or a legal guardian. Ethics approval for this study was given by the Ethics committee of the medical department of the University Marburg.

The sample consists of a total of N=33 adolescent and adult subjects (Female=2) of which N=22 received an ASD diagnosis after BEC. Additional demographic details can be seen in Table 1. The mean age of the entire sample was $18.99 (SD = 3.06)$. A series of Brunner-Munzel test (BMT from here on) (Brunner & Munzel, 2000) revealed no significant group differences for the relevant variables in Table 1 ($p > .337$) with exception of the IQ score. The IQ-score in the ASD-group ($Mdn = 95.45$) was significantly lower than in the N-ASD group ($Mdn = 114.54$) with $B = -3.60, p = .0012$.

Table 1: Demographic data of both groups with mean (M) and standard deviation (SD). Correct responses are displayed relative to all possible cards (either 96/128). Perservative errors as a subset of the overall false responses are given in proportion of the individual played cards.

	ASD (N=22)		non-ASD (N=11)	
	M	SD	M	SD
IQ score	95.45	17.99	114.54	15.84
Correct responses (%)	81.20	10.35	81.39	7.51
Persev. Err. (%)	4.26	5.98	5.29	4.83
False responses (%)	18.79	8.22	18.79	10.35

Wisconsin card-sorting task

As part of a larger diagnostic program, subjects performed a modified and computerized version of the Wisconsin Card-Sorting test (WCST) (Grant & Berg, 1948). The WCST is considered the neuropsychological gold standard in measuring executive functions like cognitive flexibility and set-shifting (Kopp, Maldonado, Scheffels, Hendel, & Lange, 2019; Lange, Seer, & Kopp, 2017) with adequate psychometric properties (Kopp, Lange, & Steinke, 2021).

In this WCST version, there were four target cards displayed on a monitor each card showing colored groups of geometric symbols. The subjects were tasked to discover a 'hidden' sorting rule by iteratively allocating one of the sorting cards per trial to one of the four target cards. After card allocation, subjects received binary feedback if the allocation followed the current sorting rule ('correct') or not ('incorrect'). There are three possible 'hidden' sorting rules i.e. sorting cards either by color, by number, or by a geometric symbol. The hidden rule changed after 10 successful allocations (completing a 'cognitive set') without explicit instructions and subjects were required to infer the new sorting rule. Note, that two versions of this computerized task were used in our data collection which had a different total numbers of sorting cards i.e. 96 and 128. However, both versions employed the same criteria for completing the task: either 6 cognitive sets of 10 successful card allocations following a current sorting rule or after all sorting cards were used.

Additional to the number of correctly allocated cards (cards with 'correct' feedback from experimenter), the WCST implementation used in this work considers four main error variables: (1) the total number of non-perseverative errors, (2)

a subset of the non-perseverative errors called the set-loss errors, (3) 'normal' non-perseverative errors and (4) perseverative errors. The total number of errors is a composite of all four error variables. Perseverative errors are defined as the continued application of the immediate prior sorting rule despite receiving at least one prior negative feedback on that sorting rule indicating a rule-change. Complementary to the perseverative errors are the non-perseverative errors and the sum over both error-types describes the total errors made with its inverse the total cards correctly allocated. Non-perseverative errors are comprised of two subsets: 'normal' errors and set-loss errors. A set-loss error describes the loss of a cognitive set i.e. applying a wrong sorting rule after initial correct identification of the change in sorting rule. A 'normal' error describes all other non-perseverative errors.
Note, that N=21 performed the 96-card version and N=12 the 128-card version. Importantly, of those performing the 128-card version only N=3 subjects required > 96 sorting cards to reach the end-criterion. To reduce the possibility of bias these subjects might contribute, all WCST error variables (Table 1) were normalized by the individual total number of played cards. A recent meta-analysis hints towards a large increase in perseverative errors with a mean cohen's D of $d = 1.47(SE = .02)$ in people with ASD when compared to mental-age matched controls (Landry & Chouinard, 2016). For brevity, only those errors are reported and discussed.

The Hierarchical Gaussian Filter

The HGF is a hierarchical probabilistic time-series model that can be understood as a hierarchical generalization of the Kalman-Filter with dynamic learning rates. The HGF was derived by C. Mathys, Daunizeau, Friston, and Stephan (2011); C. D. Mathys et al. (2014) as a generic neurocomputational model of learning under uncertainty and has been utilized multiple times in the context of ASD (Lawson et al., 2017; Sevgi, Diaconescu, Henco, Tittgemeyer, & Schilbach, 2020).

Specifically, the HGF was derived within the meta-Bayesian 'Observing-the-observer'-framework (Daunizeau, Den Ouden, Pessiglione, Kiebel, Stephan, & Friston, 2010; Daunizeau, Den Ouden, Pessiglione, Kiebel, Friston, & Stephan, 2010) i.e. this model aims to enable experimenters to infer information about (approximate) Bayesian inferences performed by a subject presented with a specific cognitive task. Specifically, the HGF-framework consists of a *generative model* and a variationally inverted *inference model*. For full mathematical derivation and assumptions see (C. Mathys et al., 2011) and (C. D. Mathys et al., 2014).

The *generative model* of the HGF is a set of formal assumption over how hidden or latent quantities (x) of a dynamic environment (e.g. cognitive task) are assumed to evolve over time. In the HGF, the environmental generative model is described as a hierarchically interconnected set of Gaussian random walks of which the lowest level produces the observed outcomes. A full treatment can be found in C. Mathys et al. (2011).
The *inference model* of the HGF consists of a *perceptual model* and a *decision model* (or response model). It describes the inverted (via Gaussian Mean-Field variational approximation) HGF generative model which can be fitted to an agent. Mean-field variational Gaussian expectations over trial-by-trial hidden environmental quantities x entertained by the agent are denoted by the $q(x)$ in Figure 1. As highlighted by C. Mathys et al. (2011); C. D. Mathys et al. (2014), this variational Bayesian inversion leads to closed-form update equations of trial-by-trial Gaussian expectations over new observations $q(\hat{x})$ akin to generic delta-learning rules. These prior expectations about sensory input are encoded by a mean $\hat{\mu}$ and precision $\hat{\pi}$ (inverse variance) for each level of the HGF and are updated via prediction errors (PE) passed up the hierarchy weighted by a dynamic learning-rate α. Once updated, prior expectations become posterior expectations (μ,π) which are then used to predict input at the next time-step $t+1$. At the lowest level, the HGF *decision model* converts prior estimates $q(\hat{x})$ over latent quantities into probabilistic expectations of observations. The HGF inference model allows binary, categorical, and continuous outcomes to be predicted. In the case of the 3-level HGF perceptual model as employed by Lawson et al. (2017) and here, the *first level* encodes trial-by-trial outcome probabilities of observations $q(x^1)$. As Lawson et al. (2017) also predicted the reaction times, their decision model was continuous. As we predicted the categories (color, geometric form, or form count), our decision model was categorical with three outcomes corresponding to the WCST sorting rules. The trial-by-trial probabilistic expectations over obseravtions are converted from the second level of the HGF via a softmax function ($s(.)$ in Figure 1). The *second level* keeps track of the unbounded outcome expectations over WCST categories $q(x^2)$ at time t. Prior expectations are parameterized by a mean $\hat{\mu}_2$ and a precision $\hat{\pi}_2$. Importantly, the precision of this expectation is a function of the trial-by-trial estimate of the next upper (third) level ($q(x^3)$) and a subject-specific model parameter ω_2 encoding the time-invariant (tonic) expected tendency of change in outcomes (volatility). Again, predictions at this level are updated via a closed-form update equation yielding posterior expectations μ_2 and π_2 used to predict the upcoming observations. Updates are calculated using the weighted prediction error ($PE_t^1 * \alpha_t^2$) from the lower level defined by the difference of the prediction $q(x_t^1)$ and observation O_t at the current trial.
As the precision $\hat{\pi}_2$ of the trial-by-trial expectation $\hat{\mu}_2$ over $q(x^2)$ is partly determined by the *third level*, this third level encodes expectations regarding the trial-by-trial (phasic) volatility of outcomes $q(x^3)$ i.e. the phasic tendency of change in the outcome expectations $q(x^2)$. Similarly, this volatility expectation is represented by a mean $\hat{\pi}_3$ and precision $\hat{\mu}_3$. Note, that the trial-by-trial precision of this expectation $\hat{\mu}_3$ is partially a function of the tonic and individually estimated model parameter called the meta-volatility estimate (ω_3). This parameter reflects the time-invariant belief of the subject in the instability of the instability in environmental contingencies. Only $\hat{\mu}_3$ is updated by a closed-form equation

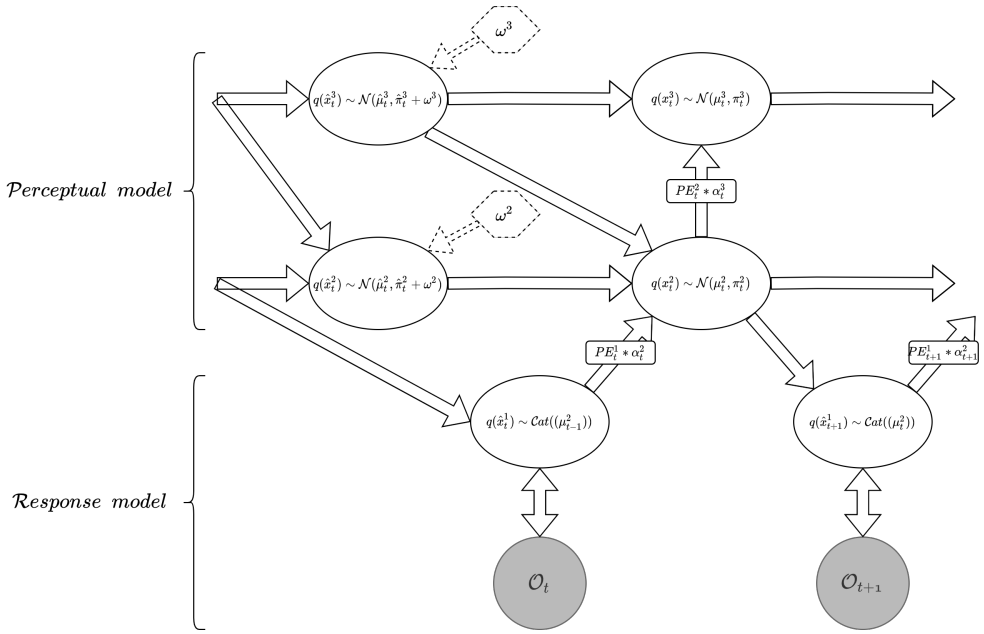

Figure 1: Schematic depiction of the three-level categorical HGF. $q(x)$ are the variational Gaussian approximations of the inverted HGF inference model over hidden states $x$as represented by the generative model of the subject. O represents input observed by the subject (and the HGF). $q(x)$ denotes trial by trial hierarchical prediction errors and α the respective weights. ω^2 and ω^3 denote individual tonic volatility estimates.

to the posterior volatility expectation μ_3.

The priors over the HGF parameters of our model were adapted from the publication of Lawson et al. (2017) when applicable. The Gaussian prior on ω_2 ($\mu = -5$ and $\sigma^2 = 16$) was adapted directly as they were estimated in an untransformed space. Note, that ω_2 was transformed after model fitting into the positive domain via exponentiation to aid in interpretation. As ω_3 was estimated in log-space in the HGF model used by Lawson et al. (2017) but in logit space here, due to differences in model derivation for the categorical HGF, the prior μ was transformed via the exponential function. The Gaussian prior for ω_3 was ($\mu = .002478$ and $\sigma^2 = 4$). We utilized the BFGS implementation for maximum likelihood estimation of the individual model parameters and starting points for the BFGS implementation were drawn from the prior distributions.The categorical HGF-model is implemented in the proprietary language used by MATLAB and was published as part of the TAPAS toolbox (Frässle et al., 2021). The model-implementation is available under an open-source license as part of the TAPAS toolbox (Frässle et al., 2021) on Github [1]. For model fitting in this study, the TAPAS toolbox was installed on a Windows 10 machine running MATLAB r2020. Model fitting of the HGF was done within a MATLAB environment accessed from a Python environment (3.7) via the MATLAB Engine API for Python. All subsequent analyses and plots were performed in Python. Supplementary material to this study (incl. code and data) can be accessed at the data_UMR Repo[2].

Original results and hypotheses

To follow up on this novel theoretical view provided by Lawson et al. (2014) on ASD, Lawson et al. (2017) tested several tenets and corollaries of the their theory in a sample of people with an ASD diagnosis (N=24) and normally developing individuals (N=25, IQ and age-matched) using a cued probabilistic learning task and a combination of computational modeling and pupillometry. More specifically, through this model-based inference Lawson et al. (2017), reported four central computational modeling results. These results may be understood as ASD-specifc computational marker of aberrant predictive processing.

Firstly, the individual model parameter ω^3 encoding meta-volatility showed a marked group difference between people with and without ASD. Specifically, this parameter showed a significant increase in people with ASD. In line with their theoretical work, the authors concluded that people with ASD have difficulties in adaptively responding to uncertainty in perception which may be manifested in this tendency to generally (over-) expect the unexpected. In other words, a maladaptive over-expectation of meta-volatility of environmental contingencies might be the manifestation of a compensatory meta-cognitive inference that one is unable to construct informative predictions regarding sensory input.

Secondly, the individual meta-volatility model parameters

[1] https://github.com/translationalneuromodeling/tapas/tree/master/HGF

[2] http://dx.doi.org/10.17192/fdr/87

made significant contributions in predicting the diagnostic status (ASD, non-ASD) of an individual.
Thirdly, Lawson et al. (2017) reported that people with and without ASD differ in their adaptations of learning rates when transitioning from stable task-phases to more volatile task-phases. Optimally when transitioning from a stable to a volatile phase, learning rates should generally increase to allow learning of recent contingencies. However, the authors reported that people with ASD adapted their weights (i.e. precision or learning-rate) on the hierarchical PE ($\Delta\alpha$) differently to the subjects without ASD. Importantly, people with ASD adapted their learning-rate less on the relevant level of outcome expectancies $q(x^2)$ in response to volatile task-phases and adapted learning rates more strongly on the level of the volatility over outcome expectancies $q(x^3)$.
Fourthly, $\Delta\alpha$ to task-phases with increased volatility of environmental contingencies were significantly predicted by individual model parameters ω^2 and ω^3. This further corroborates the first finding. Consequently, we examined the following hypotheses in this work:

1. Increased individual tonic volatility estimates ω_3 in the ASD group.

2. Prediction of the diagnostic status from individual tonic meta-volatility estimates (ω_3).

3. Decreased learning-rate modulation ($\Delta\alpha_2$, $\Delta\alpha_3$) in response to increased volatility (i.e. rule changes of the WCST). Specifically, a decreased $\Delta\alpha 2$ updating and a increased $\Delta\alpha 3$ updating in the ASD group relative to the non-ASD group.

4. Prediction of $\Delta\alpha_2$ and $\Delta\alpha_3$ by individual model parameters ω_2 and ω_3. Specifically, $\Delta\alpha_2$ should positively predicted by ω_2 and negatively predicted by ω_3. In turn, the $\Delta\alpha_3$ learning-rate modulation should be positively predicted by ω_2 and positively predicted by ω_3

Results

We will briefly summarize the results of our analyses. We used the non-parametric Median-based Brunner-Munzel test (BMT from here on) (Brunner & Munzel, 2000) for all subsequent analyses of group differences as it still allows reasonably robust inferences with heteroscedasticity, violated normality assumptions and unequal sample sizes without too much loss of statistical power (Karch, 2021; Fagerland & Sandvik, 2009). Median will be abbreviated by *Mdn* from here on. For all comparisons the common language effect-size (CLES/ Vargha and Delaney 'A'-measure) will be reported (Vargha & Delaney, 2000). The CLES is interpreted as the probability that a value from the relevant group will be greater than a value from the other group. A CLES=.5 denotes therefore stochastic equality between groups. Conventions for interpretation are provided by (Vargha & Delaney, 2000) with a small effect from 0.56 to < 0.64, a medium effect from 0.64 to < 0.71 and a large effect of 0.71.

Hypothesis 1: Group differences in tonic volatility estimates

In a first analysis, we tested for group differences of (ω_3). Our data are not in line with an increased ω_3 estimates in the ASD group ($Mdn = .0481$) compared to the NT group ($Mdn = .0538$) with $B = -.93.0, CLES = .392, p = .8444$.

Hypothesis 2: Significant prediction of diagnostic status from tonic volatility estimates

We fitted a logistic regression model using the mean-centered both ω_2 and ω_3 parameters as and the diagnostic status (ASD=1, ND=0) a the dependent variable. We utilized an l2-penalty on the coefficients and performed a Leave-one-out cross-validation procedure (LOOCV) over the prediction success. No HGF model-parameter made a statistically significant contribution in predicting diagnostic status with $z <= .025, p => .980$ for both ω_2 and ω_3. We estimated a LOOCV-corrected prediction accuracy of 66.66% using both ω_2 and ω_3 estimates as predictors.

Hypothesis 3: Group differences in updating the learning rates in response to volatile task phases

First, we tested if $\Delta\alpha_2$ was indeed decreased for the ASD group after a rule change. Between the immediate first and second trial ($\Delta\alpha_2^{t+1}$) after a rule change, we found no significantly decreased $\Delta\alpha_2^{t+1}$ for the ASD ($Mdn = .2053$) versus the NT ($Mdn = .2061$) group with $B = .076, CLES = .508, p = .530$. Similarly, we found no significantly decreased learning rate updating ($\Delta\alpha_2^{t+2}$) for the ASD ($Mdn = .377$) versus the NT ($Mdn = .224$) group with $B = -.178, CLES = .479, p = .430$ between the second and third trial after a rule change .
Secondly, we tested if $\Delta\alpha_3$ was significantly increased for the ASD group after a rule change. Between the first and second trials after a rule change ($\Delta\alpha_3^{t+1}$), we found no significant increase in updating of learning rates between the ASD group ($Mdn = .0152$) and the NT group ($Mdn = .0129$) with $B = .035, CLES = .504, p = .486$. Also, between the second and third trials after a rule change ($\Delta\alpha_3^{t+2}$), we found no significantly increased updating of learning rates between the ASD group ($Mdn = .0302$) and the NT group ($Mdn = .0213$) with $B = .588, CLES = .566, p = .281$.

Hypothesis 4: Significant prediction of learning rate adaptations by individual tonic volatility parameters

Here we assessed volatility estimate predicted $\Delta\alpha_2$ and $\Delta\alpha_3$ after a rule change. Again, we included ω_2 (and ω_3) as predictors in two separate multiple linear regression models predicting $\Delta\alpha_2$ and $\Delta\alpha_3$ for each of the two timepoints after a rule-change. The HC1 heteroscedasticity-robust covariance estimator was used (Hayes & Cai, 2007).
Using both ω_2 and ω_3 to predict ($\Delta\alpha_2^{t+1}$) returned a significant model ($F(2,30) = 19.37, P < .0001, R^2 = .316$) with ω_2

$(t = 6.19, p < .001)$ being the only significant predictor. Subsequently, when using ω_2 and ω_3 to predict $\Delta\alpha_2^{t+2}$ of the same level, the model was significant again $(F(2,30) = 3.72, P = .035, R^2 = .266)$ and ω_2 was the only significant predictor $(t = 2.72, p = .011)$. Note, that there was a large deviation from normality in this model with $(JB = 173, p < .001)$ for the Jarque-Bera test (Jarque & Bera, 1980).

Including ω_2 and ω_3 to predict $\Delta\alpha_3^{t+1}$ returned a significant model $(F(2,30) = 13.54, P < .001, R^2 = .403)$ with ω_2 again being the only driving predictor $(t = 5.19, p < .001)$. Similarly, the model predicting $\Delta\alpha_3^{t+2}$ was significant $(F(2,30) = 15.66, P < .001, R^2 = .507)$ with ω_2 again being the only driving predictor $(t = 5.46, p < .001)$.

Discussion

Following hypothesis 1, we were not able to obtain similar results for ω_3 as reported by Lawson et al. (2017). Contrary to their study, our data do not support the hypothesis of an increase in meta-volatility estimate for the ASD group. Several explanations are plausible. First, the WCST task employed here did not require the subject to perform multimodal cue-learning and had a deterministic reward schedule which might have decreased the difficulty considerably. Contrary to this, the probabilistic learning task in Lawson et al. (2017) employed an additional auditory cue with varying reliability and social stimuli. Secondly, the WCST has multiple volatile phases (i.e. rule changes) throughout the trials compared to only one volatile phase in the probabilistic learning task employed by Lawson et al. (2017). This might allow even subjects with impaired learning enough time to anticipate volatile phases. This fact likely decreased group differences in tonic meta-volatility estimates in our study. Importantly, when testing for group differences in the parameter coding the tonic volatility of changes in outcome contingencies (ω_2), we found a non-significant trend for a small effect towards an increased ω_2 for the ASD-group $(Mdn = 0.276)$ compared to the non-ASD group $(Mdn = 0.251)$ with $B = 1.40, CLES = .644, p.085$. In summary, the task environment presented to the subjects in this sample might simply be too predictable for aberrations in precision estimation to influence task-behavior.

In hypothesis 2, we found little influence of individual volatility estimates (ω_2 and ω_3) when predicting diagnostic status as overall group differences in those parameters were already small.

Following hypothesis 3, we were also not able to find similar effects for aberrations in PE weights as observed by Lawson et al. (2017). The original authors highlighted that individuals with ASD likely differed in their adaptive updating of learning rates in response to phasic volatility in the task environment i.e. transition from stable to volatile task-phase. Again, multiple explanations are possible for an unsuccessful replication. First, due to the continuous outcome of the 3-level HGF version used by Lawson et al. (2017)), they only had to test a single trial-by-trial learning rate per HGF-level between groups. The categorical HGF used here employs a parallel three-level hierarchy for each WCST sorting category which gives three learning rates per level. To still capture this effect meaningfully, we only examined the learning rates of the two WCST sorting categories that the sorting rule could change to. If e.g. the correct sorting rule was *'geometric form'* in the preceding cognitive set, we only examined the learning rates for the two remaining outcomes *'count of forms'* and *'color of forms'* as adaptations of learning-rates should manifest here most prominently. We then calculated the trial-by-trial average of the per-level learning rates over both 'learnable' WCST outcome categories giving a single average learning rate for these outcomes. This procedure in itself might have biased the inferences.

Secondly, for brevity, we opted to examine updates in learning rates up to three trials after each WCST rule-change. Note that this choice is necessarily arbitrary. Following the procedure described beforehand, this examination yielded two $\Delta\alpha$ values for each level and each WCST rule-change. Therefore, $\Delta\alpha$ for each level coded the absolute changes in learning rates for that specific level following a rule-change.

Thirdly, due to the multiple volatile phases of the WCST, we calculated the average of those per-level $\Delta\alpha$ values throughout all trials following a rule change for each subject. This yielded two average $\Delta\alpha$ values per level (t+1 and t+2) and subject. Still, there was only a marginal statistical trend of differences in individual tonic volatility parameters for ω_2. Without a meaningful group-difference in ω_3 (or ω_2) in the ASD group, there is likely also little difference in prior precision leading to little group differences in how strongly prediction errors are weighted.

Finally, regardless of time after the rule change, ω_2 appears to be the main driving parameter behind learning rate updates $\Delta\alpha_2$ and $\Delta\alpha_3$ regardless of time after rule-change. This corroborates the findings from Lawson et al. (2017) in general. In other words subjects with an increased belief in uncertainty of outcome expectations (ω_2), of which there was a marginal trend in our ASD-group, tended to update their learning rates more strongly in response to volatile task phases (WCST rule-change). Contrary to their results, we weren't able to find the negative predictive effect of ω_3 on either $\Delta\alpha_2^{t+1}$ or $\Delta\alpha_2^{t+2}$ i.e. that an increased belief in volatile volatility (i.e. meta-volatility) decreased learning rate updates on the second level of outcome expectations. Additionally, we were not able to find the positive predictive effect of ω_3 on either $\Delta\alpha_3^{t+1}$ or $\Delta\alpha_3^{t+2}$.

In summary, we were only able to find a subset of the effects reported by Lawson et al. (2017) likely due to a different task setup, different instances of a similar computational behavioral model and a small sample size. We are unable to clearly highlight which factors produced the observed data. Future research needs to examine what stimuli, structure or dynamics of a task environment are driving factors behind an overestimation of volatility in individuals with ASD.

Acknowledgments

This research was supported by the DFG Project 417284407 "Atypical perception in autism spectrum disorder: Combining computational models, functional neuroimaging and MR spectroscopy to understand aberrant perceptual mechanisms" and the "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. The authors wish to thank Gerti Gerber for assistance in data management.

References

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in cognitive sciences*, *6*(6), 248–254.

Brunner, E., & Munzel, U. (2000). The nonparametric behrens-fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *42*(1), 17–25.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181–204.

Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Friston, K. J., & Stephan, K. E. (2010). Observing the observer (ii): deciding when to decide. *PLoS one*, *5*(12), e15555.

Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Stephan, K. E., & Friston, K. J. (2010). Observing the observer (i): meta-bayesian models of learning and decision-making. *PloS one*, *5*(12), e15554.

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

Diagnostic and statistical manual of mental disorders: Dsm-5 (5. ed. ed.). (2013). Washington, DC: American Psychiatric Publishing.

Fagerland, M. W., & Sandvik, L. (2009). The wilcoxon–mann–whitney test under scrutiny. *Statistics in medicine*, *28*(10), 1487–1497.

Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., ... Stephan, K. E. (2021). Tapas: An open-source software package for translational neuromodeling and computational psychiatry. *Frontiers in psychiatry*, *12*, 680811. doi: 10.3389/fpsyt.2021.680811

Friston, K. J., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: comment on pellicano and burr. *Trends in cognitive sciences*, *17*(1), 1.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *Journal of experimental psychology*, *38*(4), 404.

Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the empathizing-systemizing theory of sex differences and the extreme male brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(48), 12152–12157. doi: 10.1073/pnas.1811032115

Haker, H., Schneebeli, M., & Stephan, K. E. (2016). Can bayesian theories of autism spectrum disorder help improve clinical practice? *Frontiers in psychiatry*, *7*, 107. doi: 10.3389/fpsyt.2016.00107

Happé, F. (2005). The weak central coherence account of autism. In F. R. Volkmar (Ed.), *Diagnosis, development, neurobiology, and behavior* (Vol. 21, pp. 640–649). Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9780470939345.ch24

Happé, F., & Frith, U. (2020). Annual research review: Looking back to look forward - changes in the concept of autism and implications for future research. *Journal of child psychology and psychiatry, and allied disciplines*, *61*(3), 218–232. doi: 10.1111/jcpp.13176

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in ols regression: An introduction and software implementation. *Behavior research methods*, *39*(4), 709–722.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, *6*(3), 255–259.

Karch, J. D. (2021). Psychologists should use brunner-munzel's instead of mann-whitney's u test as the default nonparametric procedure. *Advances in Methods and Practices in Psychological Science*, *4*(2), 2515245921999602.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron*, *100*(2), 424–435.

Knill, D. C., & Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712–719.

Kopp, B., Lange, F., & Steinke, A. (2021). The reliability of the wisconsin card sorting test in clinical practice. *Assessment*, *28*(1), 248–263. doi: 10.1177/1073191119866257

Kopp, B., Maldonado, N., Scheffels, J. F., Hendel, M., & Lange, F. (2019). A meta-analysis of relationships between measures of wisconsin card sorting and intelligence. *Brain sciences*, *9*(12). doi: 10.3390/brainsci9120349

Landry, O., & Chouinard, P. A. (2016). Why we should study the broader autism phenotype in typically developing populations. *Journal of Cognition and Development*, *17*(4), 584–595. doi: 10.1080/15248372.2016.1200046

Lange, F., Seer, C., & Kopp, B. (2017). Cognitive flexibility in neurological disorders: Cognitive components and event-related potentials. *Neuroscience and biobehavioral reviews*, *83*, 496–507. doi: 10.1016/j.neubiorev.2017.09.011

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature neuroscience*, *20*(9), 1293–1299. doi: 10.1038/nn.4615

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in human neuro-*

science, *8*, 302. doi: 10.3389/fnhum.2014.00302

Lord, C., & Bishop, S. L. (2015). Recent advances in autism research as reflected in dsm-5 criteria for autism spectrum disorder. *Annual review of clinical psychology*, *11*, 53–70. doi: 10.1146/annurev-clinpsy-032814-112745

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience*, *5*, 39. doi: 10.3389/fnhum.2011.00039

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the hierarchical gaussian filter. *Frontiers in human neuroscience*, *8*, 825. doi: 10.3389/fnhum.2014.00825

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of autism and developmental disorders*, *36*(1), 27–43. doi: 10.1007/s10803-005-0040-7

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. doi: 10.1016/j.tics.2012.08.009

Perrykkad, K., & Hohwy, J. (2020). Modelling me, modelling you: the autistic self. *Review Journal of Autism and Developmental Disorders*, *7*(1), 1–31.

Sevgi, M., Diaconescu, A. O., Henco, L., Tittgemeyer, M., & Schilbach, L. (2020). Social bayes: using bayesian modeling to study autistic trait–related differences in social cognition. *Biological Psychiatry*, *87*(2), 185–193.

Tewes, U., Rossmann, P., & Schallberger, U. (1999). Hamburg-wechsler-intelligenztest für kinder - 3. auflage.

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological review*, *121*(4), 649–675. doi: 10.1037/a0037665

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the cl common language effect size statistics of mcgraw and wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101–132.

Waye, M. M. Y., & Cheng, H. Y. (2018). Genetics and epigenetics of autism: A review. *Psychiatry and clinical neurosciences*, *72*(4), 228–244. doi: 10.1111/pcn.12606