An Adaptive Signal Detection Model Applied to Perceptual Learning

Percy K. Mistry[1] **(pkmistry@uci.edu)**
Joshua Skewes[2] **(filjcs@cas.au.dk)**
Michael D. Lee[1] **(mdlee@uci.edu)**
[1]Department of Cognitive Sciences, University of California Irvine, Irvine, CA 92697-5100 USA
[2]Interacting Minds Centre, Department of Culture and Society, Aarhus University, Denmark

Abstract

We introduce a new model of adaptive criterion setting within a signal detection framework, and show how this provides psychological insights that allow us to segregate causes of suboptimality in perceptual learning. We apply this to a perceptual learning task for both neurotypical and autistic participants. The model parameters provide a bridge between the mechanisms of an aberrant precision account of autism and resulting behavior that can be interpreted within a receiver operating characteristic framework. The model makes superior out-of-sample predictions compared to standard signal detection theory, about how people adapt to different environmental manipulations when asked to categorize audio-spatial stimuli. We find that accuracy of participants is more strongly correlated to the construct of persistence signals that inhibit response flexibility, than to the neuromodulatory gain. We also find evidence for individual differences in persistence that are correlated to scores on the autistic traits questionnaire.

Keywords: adaptive signal detection, autism, cognitive model, categorization

Introduction

Autism spectrum disorder (ASD) is a highly prevalent condition with about 1 in 68 affected globally. Sensory symptoms are common in ASD, and include hypo- and hypersensitivity to stimulus, and sub-optimalities in perceptual inference (Turi et al., 2015). One area in which perceptual differences are particularly common in autism is auditory perception (OConnor, 2012), including auditory localization (Teder-Sälejärvi, Pierce, Courchesne, & Hillyard, 2005). Skewes and Gebauer (2016) examined the potential cause of suboptimality in perceptual judgments for the spatial sources of sounds in adults with ASD.

Classical signal detection theory (SDT) analysis showed that both ASD and neurotypical (NT) participants did adapt their criteria in response to base rate and discriminability manipulations, but did so suboptimally. Adults with ASD had a larger deviation from optimal categorization of stimuli than NT participants. On average, ASD participants showed lower discriminability and less extreme criterion setting, although these differences were not statistically significant. Classical SDT analysis can account for behavioral patterns extremely well within a fixed environment. It does not, however, provide a descriptive account of how people adapt their criterion in response to environmental manipulations, such as changing base rates, changing discriminability, or changing utilities for different types of correct decisions and errors (although it prescribes what the normative change in criteria might be). Given the limitations of SDT analysis, it is not clear whether the observed behavioral differences are on account of differences in sensory precision, in contextual integration of prior expectations, or due to differences in executive functioning. The latter may manifest as differences in the flexibility with which response strategies are changed as feedback changes. There are some existing theories of how people may adopt a flexible rather than static criterion across trials (Treisman & Williams, 1984; Erev, 1998; Turner, Van Zandt, & Brown, 2011). In this paper we introduce an alternative adaptive account of how people set criteria for categorizing stimuli. We show that our new model has a direct interpretation both within the ROC framework of classical SDT and within the aberrant precision account of ASD (Lawson, Rees, & Friston, 2014). This helps shed light on the potential causes of differences in suboptimality that are observed in ASD and NT participants. It also allows us to improve the predictive capability for how people might adapt their criterion in different environments.

Experimental Data

Our data come from experiments reported by Skewes and Gebauer (2016), in which, on each trial, participants had to categorize an auditory stimulus into one of two categories. The categorization was based on a cover story of classifying different species of crickets, with the territory of one species being distributed to the left and the other to the right. Based on the spatial location of the sound stimulus, participants had to categorize which species the sound on each trial originated from. Each trial was followed by corrective feedback. The stimuli for the two species were spatially overlapping to some extent to introduce uncertainty into the task. Each participant completed 960 trials split into 4 randomized blocks. The 4 blocks consisted of a 2 X 2 factorial design, with each block having either a low (25%) or high (75%) base rate (BR) of one species, and a low or high discriminability. The blocks were presented in randomized order. In the low discriminability environment there was greater spatial overlap of the auditory stimulus from the 2 species.

The key results from a classical SDT analysis were that both ASD and NT participants showed sensitivity to base rate as well as discriminability manipulations. This sensitivity, however, was suboptimal, and both groups demonstrated significant deviation from the optimal response criterion, as shown in Figure 1. This deviation was larger for the ASD group than for the NT group for all 4 conditions. As a result ASD participants also demonstrated lower accuracy as shown in Figure 2. A one-sided Bayesian t-test (JASP-Team, 2016) produced a Bayes factor (BF) of 4.0 in favor of the accuracy for ASD participants (mean 73.7%) being lower than

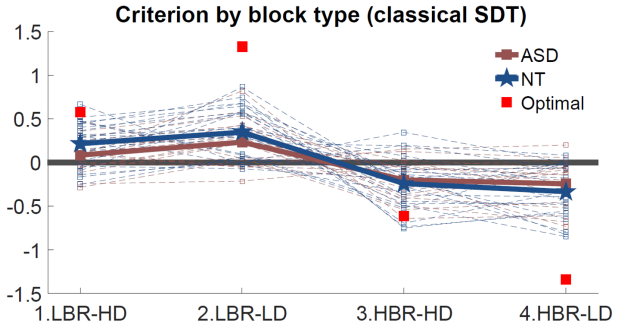

Figure 1: Inferred criterion based on classical SDT for individuals (dotted lines) and group means (thick lines) in the 4 blocks that vary in base rate (LBR=low; HBR=high) and discriminability (LD=low; HD=high). The red squares show the optimal criterion placement.

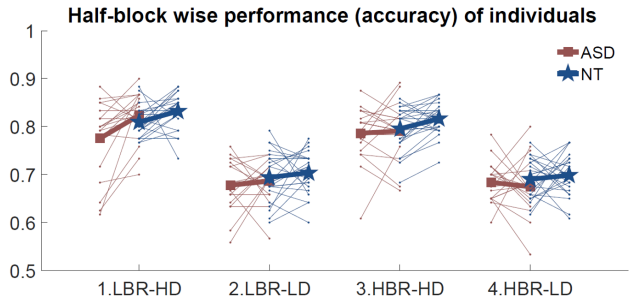

Figure 2: Accuracy of categorization for individuals (dotted lines) and group means (thick lines) in the 4 types of blocks. The blocks are split into two halves of 120 trials each, so the slope of the lines shows within block changes. The NT and ASD plots are displaced adjacent to each other to improve the clarity of the figure. ASD participants show greater variability and some show lower levels of performance, but the differences at a group level are very small.

NT participants (75.5%). ASD participants demonstrated less extreme criterion values in response to base rate manipulations, but the BF for this was not conclusive. Figure 2 also shows the performance of individual participants divided into the first 120 and second 120 trials, for each of the 4 types of blocks. There does not seem to be a significant improvement within blocks for either group. In general, accuracy was lower in conditions with lower discriminability. The participants in the ASD group show greater variability, especially in the lower accuracy range. In the following sections, we introduce a new framework for adaptive criterion setting, and then analyze data from 19 ASD and 23 NT participants (this included all the participants recruited for the experiment except 1 participant with ASD who did not complete the experiment).

Adaptive criterion setting

For this task, the criterion is defined as the spatial boundary such that any stimulus perceived to come from the right of this criterion is categorized as species 1 and from the left as species 2. As a matter of convention, objective spatial locations to the right are given positive values and to the left are given negative values, so that the species on the right is considered the "signal" and on the left, the "noise". On the tth trial, participants are assumed to adapt a criterion c_t, such that if their perceived stimulus m_t is higher than c_t, they identify the stimulus as species 1, and if $m_t < c_t$, they identify the stimulus as species 2. Our adaptive SDT model (ASDT) assumes that people do not adapt a fixed criterion across all trials, but keep changing the criterion in response to feedback. Such changes would be responsive to differences in rewards, the perceived size of the error, or the past history of correct and incorrect feedback. In this task rewards are symmetric, that is, there is no difference in the rewards for correctly identifying species 1 (hit) or species 2 (correct rejection). Similarly there is no difference in the penalty depending on whether species 1 (miss) or species 2 (false alarm) was incorrectly identified. Since the two categories were fictional, there is no reason to believe that participants have an inherent bias towards either. Accordingly, ASDT groups all correct and all incorrect decisions together. It is assumed that people shift their criterion only after receiving feedback about errors, but see the discussion section for possible counterarguments.

Formally, ASDT assumes that:

$$c_t = c_{t-1} + I_{W_{t-1}} \left(\frac{(\eta_t + \rho_t)}{1 + \Sigma_{i=1}^{t-1} \alpha^{t-i}} \right) \quad (0 \le \alpha \le 1) \qquad (1)$$

$$\eta_t = \delta(m_{t-1} - c_{t-1}) \quad (0 < \delta \le 1) \qquad (2)$$

$$\rho_t = \delta \left(\Sigma_{i=1}^{t-1} \{ \alpha^{t-i} (m_i - c_i) \} \right) \qquad (3)$$

Here $c_1 = 0$, and δ, α are individual level parameters: δ represents gain control and α represents persistence, or the lack of response flexibility. $I_{W_{t-1}}$ is an indicator function that is 1 if the $(t-1)$th trial was incorrect and 0 otherwise. The term $(m_{t-1} - c_{t-1})$ is the underlying difference signal between the stimulus and criterion, and represents the error signal on incorrect trials. If the $(t-1)$th trial was a miss because the criterion was too high, this term will be negative and serve to lower the criterion. If the previous trial was a false alarm because the criterion was too low, this term will be positive and serve to increase the criterion. This difference signal is modulated by a gain control parameter δ. Higher values of δ imply a larger corrective feedback given a particular level of sensory feedback. The resulting term η_t is the contribution of the immediately preceding trial to the corrective feedback, which we call the *immediate signal*.

Note that if the previous trial is a hit or correct rejection, the criterion will not change. However, if the previous trial is incorrect, the change made includes feedback based not only on the immediately preceding trial, but also feedback weighted and averaged from previously experienced trials, including correct trials. On correct trials, the difference signal term is positive for hits and negative for correct rejections. The cumulative contribution of all previous trials is given by the ρ_t term, which we term *persistence signal*. Here, the weight given to older feedback keeps decreasing, and is a function of α. The feedback term from j trials earlier is given a weight

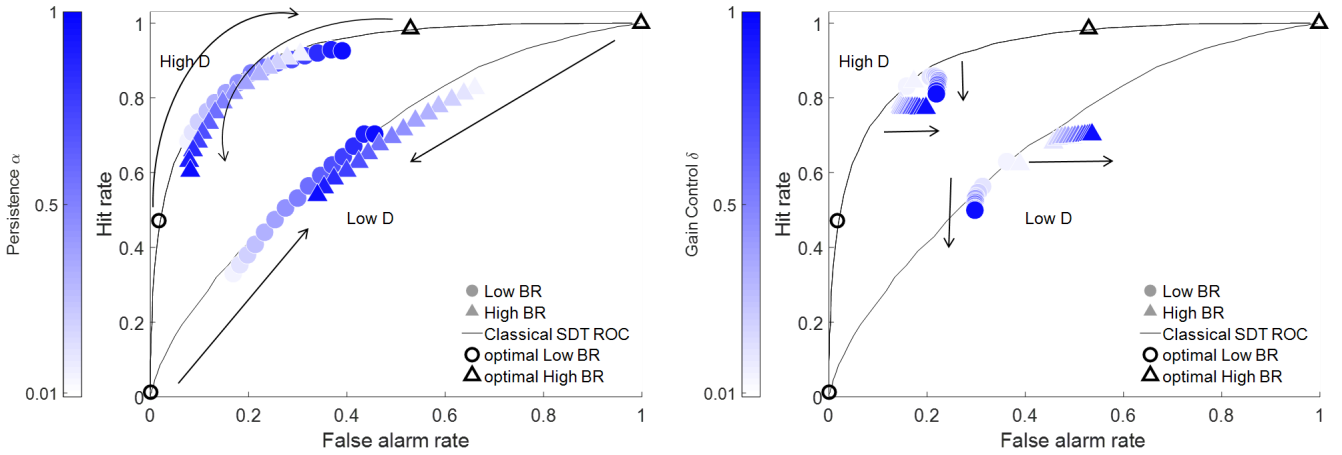

Figure 3: ROC curves for high-D and low-D environments (i.e. different experimental level of discriminability), and the hit rate and false alarm rate based on optimal criterion placement for both low BR and high BR conditions. The colored plots show how a change in α (left panel) and δ (right panel) affect how individual behavior moves away from optimality. Increasing α results in movement along the ROC (does not affect sensitivity or discriminability), but changes in δ shift performance to a lower ROC (impact sensitivity).

of α^j. A weighted average of η_t and ρ_t is then computed as the final corrective feedback for the criterion level. Having large values of α result in a weighted average over a longer time window, leading to higher persistence of sensory feedback and lower flexibility of response changes. Since α acts as a discount rate for previously acquired feedback, a value of $\alpha = 1$ means that on each trial, the effective feedback is the mean value of feedback acquired on all trials experienced so far. A low value of α, close to 0, would mean that feedback from only the most recent trial is taken into account.

Classical SDT often uses receiver operating characteristic (ROC) analysis which plots performance in terms of hit and false alarm rates. Figure 3 characterizes ASDT in ROC terms, showing the results of simulations that systematically varied α and δ in 4 different environments that varied in terms of base rate and discriminability, similar to the experimental paradigm. We show that ASDT has a strong relationship to the dynamics of the ROC. The resulting performance is plotted along the ROC in figure 3. The left panel shows the sensitivity to α. As α increases from 0.01 to 1, it results in a smooth change *along the ROC*, and *away* from the optimal criterion. For very low values of α, behavior still deviates from optimal performance, but to a smaller extent. Very high values of α close to 1 show maximum deviation away from optimality. The right panel of Figure 3 shows the sensitivity to δ, with increasing values of δ showing *a movement away from the ROC*, with reducing hit rates in low BR and increasing false alarm rates in the high BR conditions. Thus α and δ capture two separable behavioral deviations: along the ROC or away from the ROC. Changes in δ capture what in traditional SDT analysis, is captured as a difference in sensitivity. We note that the simulations show that the mean criterion level is extremely sensitive to values of α, with higher values of α leading to less extreme average criterion values. Higher values of δ result in higher δ in the criterion across trials.

Modeling Results

Model description

Figure 4 shows the stimulus and criterion based on classical SDT as well as our adaptive SDT model for a single participant. The effectiveness of adaptive SDT is especially visible in the predictions in the low discriminability (LD) blocks. Figure 5 shows the application of our adaptive SDT model to 2 NT and 2 ASD participants. The achieved accuracy of the 4 participants and the mean values of the inferred α and δ parameters for these participants are shown on the left side. The 4 participants were selected to show behavior where both parameter are low (participant 1), low α but high δ (participant 2), high α and low δ (participant 3), and both parameters high (participant 4). The first column shows the immediate sensory error signature η_t across all 960 trials. It can be seen that participants 2 and 4, with higher values of δ, show higher η values. The second column shows the persistence related error signature ρ_t, and here, participants 3 and 4, with high values of α, show higher ρ_t values. The third column shows the sum of these two, which is what contributes to the total criterion correction on each trial. Of interest is the fact that across most trials, the persistence based feedback signature seems to be the inverse of the immediate sensory error feedback, thus leading to muted corrections when α is large. The last column shows the resulting criterion movement from trial to trial. All four participants show some sensitivity to BR and SD, but this is much higher in participants 1 and 2, who accordingly show higher accuracy rates.

Inference about individual parameters

We then use the complete data set to infer individual level α and δ parameters for the 19 ASD and 23 NT participants. Figure 6 shows the joint posterior density of the parameters for the two groups. The size of the squares is the joint probability density. The overall densities look quite similar for ASD and

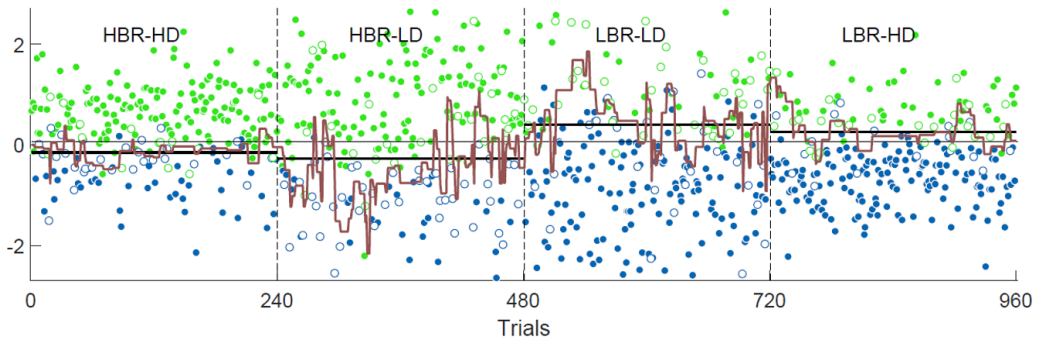

Figure 4: Criterion dynamics for participant 2: The brown line shows the inferred adaptive criterion inferred by ASDT. The thin black line shows the criterion based classical SDT, combining all 960 trials together. The thick black lines in each block lines show the criterion based on classical SDT computed for each block separately. The dots show the standardized stimulus values. Filled green dots show hits, filled blue dots show correct rejections. Empty green dots show false alarms and empty blue ones show misses. A model that predicts well should show green dots above the criterion, and blue dots below the criterion.

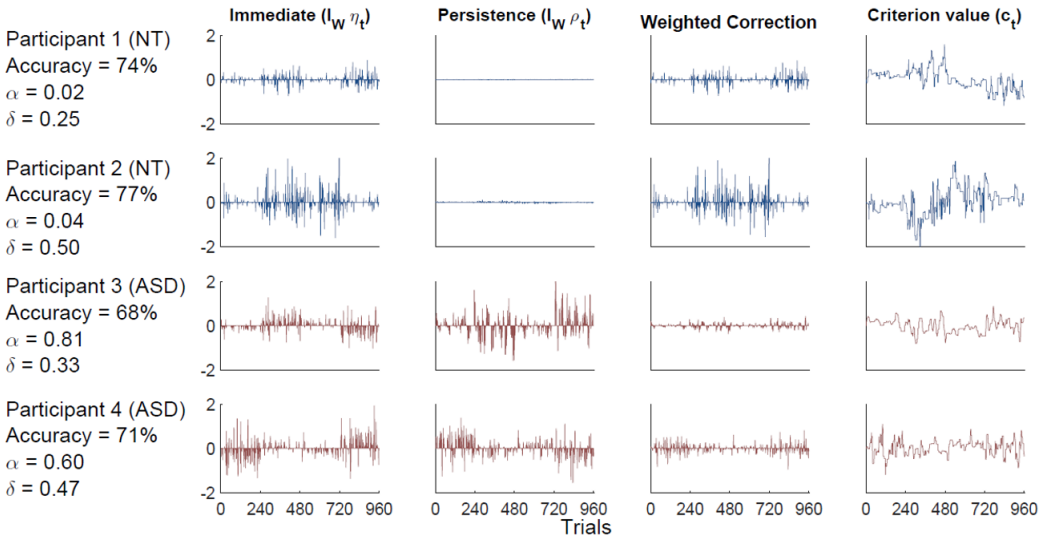

Figure 5: A process perspective inferred from the model for 4 of the 42 participants, to show how the adaptive model infers distinct forms of behavior. The four columns show η_t, ρ_t, $\eta_t + \rho_t$, and c_t.

NT participants. The ASD group shows slightly higher values of α (mean 0.26, SD 0.26) versus the NT group (mean 0.19, SD 0.20), and similar values for δ (mean 0.4, SD 0.09 for ASD versus mean 0.41, SD 0.07 for NT) but neither is significant. A Bayesian t-test suggests no main effect of diagnosis (ASD vs NT) on either parameter with BFs of 0.47 and 0.31 respectively, testing for a difference between the two groups for α and δ. A Bayesian ANOVA analysis however reveals a significant main effect of the Autistic traits questionnaire (AQ) score, with a Bayes factor of 4.4. Higher AQ scores demonstrate higher values of α. In Figure 6 the color represents the weighted AQ scores. For NT participants, this score is almost uniformly low as expected, except for the highest level so fα within NT participants. With a few exceptions, α seems to increase with an increasing mean AQ score, shown by the density clusters in dark red towards the right.

A Bayesian test of correlation yields strong evidence for a negative correlation between α and the actual accuracy of participants in the task ($r = -0.54, BF154$), and mild evidence for a positive correlation between δ and accuracy ($r = 0.39, BF = 4$). This supports the notion that any suboptimality is driven primarily by higher persistence signal (α), than by the gain (δ).

Model performance

We implemented ASDT within a Bayesian inference framework for statistical inference (Plummer et al., 2003). To test the model, we infer the parameters using only data from one of the blocks at a time and calculate the accuracy of the out-of-sample predictions for the remaining 3 blocks based on the mean posterior predictives. A floor benchmark is the accuracy with which the classical SDT based criterion calculated using the hit rate and false alarm rate from a single block is able to predict the responses for the remaining blocks. Table 1 shows a comparison of the predictions based on using data

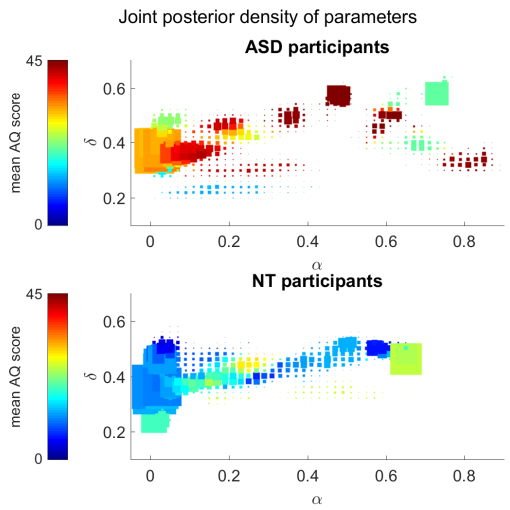

Figure 6: The joint posterior probability densities for the 2 model parameters for the ASD and NT participants. The size of the squares shows the probability density and the color shows the posterior probability weighted mean AQ (autistic traits questionnaire) score for that particular combination of α and δ values.

from each of the 4 blocks for the classical SDT and ASDT models. The ASDT model provides superior predictions, and provides a psychological process perspective to explain how the criterion adapts over time.

Table 1: Accuracy of out-of-sample predictions using the difference blocks (LB=Low base rate; HB=High base rate; LD=Low discriminability; HD=High discriminability). SDT is based on classical SDT analysis, and ASDT is based on our proposed model of adaptive criterion setting.

	Out of sample prediction using block				
	LB-HD	LB-LD	HB-HD	HB-LD	All
Autism Spectrum Disorder (ASD)					
SDT	81.2%	82.0%	80.8%	79.9%	81.0%
ASDT	85.8%	86.0%	85.9%	86.1%	86.0%
Neurotypical (NT)					
SDT	82.1%	78.8%	78.9%	79.7%	79.9%
ASDT	87.0%	87.1%	87.3%	86.6%	87.0%

Aberrant precision interpretation

Suboptimality in sensory (and other) tasks by adults with ASD has been proposed to be a disorder of metacognition (Friston, Lawson, & Frith, 2013; Van de Cruys et al., 2014). Within this framework, Lawson et al. (2014) propose two mechanisms that constitute an aberrant precision account of autism. The first is enhanced neuromodulatory gain for how prediction errors are encoded in individuals with autism. Adaptive gain control in neurotypical individuals is expected to adjust to environmental volatility so that there is higher gain in more volatile environments. It has been proposed that in individuals with autism however, gain control might be excessively enhanced because of the expectation of highly precise sensory inputs. This in turn would lead to a lack of context sensitivity, as reported by Palmer, Paton, Kirkovski, Enticott, and Hohwy (2015). Thus we conclude that the gain control processes controlled by δ in our model corresponds to this mechanism. We would thus expect to see higher values of δ for ASD participants under this framework. We do not however observe this and δ values for both groups are strikingly similar. We propose that excess neuromodulatory gain control is not a key driver of suboptimality for the Skewes and Gebauer (2016) task. This result supports the conclusion that autism is not characterized by uniform differences in the weighting of prediction error (Manning, Kilner, Neil, Karaminis, & Pellicano, 2016).

The second mechanism under the predictive coding framework constitutes a lack of sensory attenuation, sometimes manifested as a failure to suppress prediction errors generated by repetitive stimuli over time (e.g. Kleinhans et al. (2009)), or in failing to notice changes in the predictive value of specific information (Van de Cruys et al., 2014). The key aspect is that individuals with autism can form accurate representations of low-level prediction errors, but the translation of these into higher level signals differs when compared to NT individuals. Specifically, the higher level signals might be influenced to drive repetitive behavior and perceive prediction errors over time in a consistent manner. This may thus lead to behavior that is more resistant to change. In our model, we propose that α captures this mechanism. High values of α would indicate persistence of sensory feedback over time, leading to increased consistency of actions and longer time frames to respond to environmental changes. We would expect to see higher values of α for ASD participants under this framework. We see some indication of this, as values of α do show a small but significant increase with increasing AQ scores. We propose that increased persistence and thus a lack of response flexibility is the key driver for any increased suboptimality observed in this pool of ASD participants. Relating this to classical SDT analysis, increased lack of response flexibility would result in an increase in deviation *along* the ROC, not necessarily demonstrating reduced sensitivity.

There is general consensus that lower level sensory error signals can be more precise, but are transformed into attenuated or less precise higher level prediction error signals in people with ASD. A perspective for explaining this has been using Bayesian updating (Pellicano & Burr, 2012). The basic idea is that individuals with ASD may demonstrate inefficient Bayesian updating since they may have diffused priors, called hypo-priors, but strong sensory signals. We propose a related but slightly different explanation. Even if individuals with ASD start with diffused priors, updating with a strong sensory signal on a trial by trial basis would result in sharp posteriors. Since the posterior on one trial would form the basis for the prior on the next, a diffused prior would not be sustainable over trials. A sustained diffused prior might however be maintained from trial to trial if apart from a strong sensory signal, there was a second signal that also influenced these

priors. On any trial, if all previous error information has been accounted for efficiently in the updated prior, Bayesian updating would require that only new information is taken into account for further changes to be made to the criterion. This is represented by the term η_t. Hence any significant contribution from ρ_t leads to interference and ineffective updating. Even if η is a sharp sensory signal, if ρ is partly in opposition to η, the result would result in sustained diffused beliefs, as have been proposed in theory. Slightly higher levels of α and the resulting higher values of the ratio of absolute magnitudes of ρ_t to η_t (mean ratio of 5.4 for ASD versus 2.9 for NT) though not statistically significant, directionally align with Pellicano and Burr (2012), who suggest that autistic perception might suffer from hypo-priors.

Conclusion

We have developed and applied an adaptive SDT model that can provide additional psychological insight and help in segregating causes of sub-optimality in perceptual learning. The three primary contributions of this work are:

1. Providing a successful account of sequential effects in perceptual judgment
2. Model based evidence of differences (or the lack thereof) in inferred parameters between the ASD and NT groups.
3. Implications of this model-based account for interpreting the aberrant precision hypothesis.

We present statistical evidence that persistence signals have a stronger impact on suboptimality than neuromodulatory gain. However there is only mild evidence of any differences between the neurotypical and ASD participants in their optimality or underlying persistence levels.

An attractive feature of ASDT is that the same two proposed parameters have direct interpretations both within the framework of classical SDT (separability in behavior along the ROC vs off the ROC) and within the prevailing aberrant precision account of ASD (persistence and gain control). It also provides superior out-of-sample predictions.

There are two key limitations that need to be tackled in future work. The first is the assumption that there is no updating that takes place after a correct trial. It is plausible that correct trials provide confirmatory feedback based on which individuals might become more risk-seeking with their criterion setting. Secondly, the model does not incorporate asymmetric utilities for different types of correct and incorrect responses. Future work could include a reward utility based adjustment that allows correction to be skewed in a particular direction because of objective or perceived skews in the rewards and penalties.

References

Erev, I. (1998). Signal detection by human observers: a cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological review*, *105*.

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

JASP-Team. (2016). Jasp (version 0.8.0.0)[computer software].

Kleinhans, N. M., Johnson, L. C., Richards, T., Mahurin, R., Greenson, J., Dawson, G., & Aylward, E. (2009). Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *American Journal of Psychiatry*, *166*.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in human neuroscience*, *8*.

Manning, C., Kilner, J., Neil, L., Karaminis, T., & Pellicano, E. (2016). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science*.

OConnor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience & Biobehavioral reviews*, *36*.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G., & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: a predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*.

Pellicano, E., & Burr, D. (2012). When the world becomes too real: a bayesian explanation of autistic perception. *Trends in cognitive sciences*, *16*.

Plummer, M., et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124).

Skewes, J. C., & Gebauer, L. (2016). Brief report: suboptimal auditory localization in autism spectrum disorder: support for the bayesian account of sensory symptoms. *Journal of autism and developmental disorders*, *46*.

Teder-Sälejärvi, W. A., Pierce, K. L., Courchesne, E., & Hillyard, S. A. (2005). Auditory spatial localization and attention deficits in autistic adults. *Cognitive Brain Research*, *23*.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological review*, *91*.

Turi, M., Burr, D. C., Igliozzi, R., Aagten-Murphy, D., Muratori, F., & Pellicano, E. (2015). Children with autism spectrum disorder show reduced adaptation to number. *Proceedings of the National Academy of Sciences*, *112*.

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological review*, *118*.

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*.