**Analysis of perceptual bias reveals slow updating in autism and fast forgetting in dyslexia**

Itay Lieder[1†], Vincent Adam[2†], Or Frenkel[4], Sagi Jaffe-Dax[3], Maneesh Sahani[2†] & Merav Ahissar[1,4*†]

**Affiliations:**

[1]Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem.

[2]Gatsby Computational Neuroscience Unit, 25 Howland Street, London, W1T 4JG.

[3]Psychology Department, Princeton University, Princeton NJ.

[4]Psychology Department, Hebrew University.

[†]Equal contribution

*Correspondence to: msmerava@gmail.com

**Abstract:**

Individuals with autism and individuals with dyslexia both show reduced use of previous sensory information (stimuli statistics) in perceptual tasks, even though these are very different neurodevelopmental disorders. To better understand how past sensory information influences the perceptual experience in these disorders, we first investigated the trial-by-trial performance of neurotypical participants in a serial discrimination task. Neurotypicals over-weighted recent stimuli, enabling fast updating of internal sensory models, which is adaptive in changing environments; they weighted longer-term stimuli statistics according to the detailed stimuli distribution, enabling reliable predictions of upcoming stimuli, which is adaptive in stable environments. Compared to neurotypicals, individuals with dyslexia weighted longer-term statistics less heavily, whereas individuals with ASD weighted recent stimuli less heavily. Thus, investigating the dynamics of perceptual inference reveals that individuals with dyslexia rely more on information about the immediate past, whereas perception in individuals with ASD is dominated by longer-term statistics.

Introspection suggests that perception merely mediates reliable responses to incoming stimuli. Yet sensory input is typically noisy and ambiguous. Our rich and coherent perception

is the outcome of integrating our sensations with our internal models, which reflect our estimates of the environment's stimuli statistics. These models are updated and enriched with experience and, following substantial practice, support quick and effortless perception of a broad range of familiar stimuli.

Recent findings suggest that these implicit inference processes may be altered in individuals with neurodevelopmental disorders, and that their ability to use sensory statistics to update their internal models may be impaired. For example, it has been suggested that individuals with developmental dyslexia, who remain slow readers even after years of practice[1], are less efficient in learning the statistics of serially presented stimuli[2–5]. This difficulty is expected to reduce the efficiency of their predictions in a broad range of (speech[6] and non-speech[7]) stimuli. Importantly, it is expected to reduce their benefits from language statistics (e.g. syllable distribution, morphology), which typically substantially facilitates reading rate and accuracy. Indeed, this deficiency has been proposed as a core impediment to attaining expert reading performance[2-5]. Interestingly, reduced integration of stimuli statistics ("hypo-priors"[8]) in several perceptual tasks was also observed for individuals with autism (also termed ASD, Autism Spectrum Disorder). These observations have been used to explain well-documented phenomena in ASD, including reduced adaptation and the often reported feeling of being overwhelmed with sensory stimuli[9-11].

The similarity between ASD and dyslexia in underuse of prior perceptual information is puzzling since these two populations have substantially different behavioral strengths and weaknesses. While individuals with dyslexia fail to become expert readers but are not socially challenged, high-functioning individuals with ASD who have no language deficits, have difficulties deciphering social cues[12] but have no difficulty in decoding written script. In

fact, in rare cases their decoding skills may even be superior (hyperlexia[13]) to those of the general population.

In this study we set out to examine the differences between neurotypicals, individuals with dyslexia and individuals with ASD in the acquisition and use of prior perceptual information. Administering a serial 2-tone frequency discrimination task allowed us to examine the impact of (i.e. bias by) previous trials using simple yet unfamiliar (pure tones are not ecological) stimuli, and quantify separately the contribution of very recent compared with earlier trials. We found that, compared to matched neurotypical groups, individuals with dyslexia rely more heavily on the immediate past, while perceptual bias in individuals with ASD is dominated by longer-term statistics. To better understand the balance between the effects of recent and longer-term statistics, we administered different stimuli distributions to large neurotypical populations. Using non-linear analyses we found that their use of recent statistics is distribution invariant, whereas their use of longer-term statistics reflects optimal (Ideal Observer's) integration of stimuli distribution. Applying this analysis to both individuals with dyslexia and individuals with ASD revealed a double dissociation. Individuals with ASD showed reduced integration of recent statistics whereas recent effect in neurotypicals and individuals with dyslexia were similar, both in its magnitude and pattern. In contrast, the integration of longer-term statistics in individuals with dyslexia was partial and sub-optimal whereas integration of longer-term statistics in individuals with ASD resembled that of an Ideal Observer (IO). These differences in perceptual inferences parallel the slow yet successful formation of high-resolution categorical representations in ASD[14], and the fast formation of lower resolution categorical representations in dyslexia[7].

## Results

### The impact of sensory history is reduced in both ASD and dyslexia.

Contraction bias — the general tendency to estimate small stimuli as larger and large stimuli as smaller — is a well-known phenomenon[15–17] that reflects the influence of environments' statistics on perception. In natural environments it allows the formation of the most likely percept even when sensory sampling is noisy. In laboratory conditions we can use this bias as a window into implicit processes of integrating stimuli statistics to perception. We now studied this bias using serial 2-tone frequency discrimination (illustrated in **Fig. 1a**), where performers tend to "contract" the first stimulus in each trial towards the mean value of the preceding stimuli (**Fig. 1b**). Such contraction integrates the responses to the current stimulus with information on stimuli statistics based on previous stimuli (e.g. their mean). The importance of such information increases when responses are noisier. The enhanced contraction of the first stimulus in each trial of delayed discrimination is interpreted (within the Bayesian framework) as resulting from its noisier representation compared with that of the second stimulus, due to the task-imposed need to encode it in memory and retain it till the presentation of the second stimulus[16].

The impact of contraction bias (of the 1$^{st}$ tone) on trial performance depends on the relations between the first and second stimulus of the trial with respect to previous stimuli. If the first stimulus is closer to the mean than the second stimulus (Bias+ trials; **Fig. 1b**, left), its contraction increases the perceived distance between the stimuli, and consequently increases accuracy. Conversely, in trials where the first stimulus is farther from the mean (Bias− trials; **Fig. 1b**, right), contraction of the first stimulus towards this mean decreases the perceived

difference and the success rate is reduced. Consequently, participants make fewer errors when the first stimulus is closer to the mean of previous trials than the second stimulus (**Fig. 1b**). This history-driven difference in accuracy between these two types of trials can be very large (e.g., chance-level performance in Bias− trials and ~90% correct in Bias+ trials[18-20]). The perceptual bias to the mean is directly reflected in the bias of participants' responses. They tend to answer that the 2nd stimulus is higher ("$f_2 > f_1$"; **Fig. 1b** upper region) when $f_1$ is above the mean and that the 1st is higher ("$f_1 > f_2$"; **Fig. 1b** lower region) when $f_1$ is below the mean. This bias can be integrated into the common description of behavior in discrimination tasks, the psychometric function $p(\text{"}f_2^t > f_1^t\text{"}) = \Phi(\alpha\delta^t + b_0)$, where Φ, is the standard normal cumulative distribution function (i.e., $\Phi^{-1}$ is the probit function), $\delta^t$ is the veridical difference between the tones in trial $t$, $\alpha$ is the individual sensitivity to this difference, and $b_0$ is the bias to perceive $f_2^t$ as higher. Contraction bias can be described with respect to the psychometric function by introducing a trial-by-trial shift of the psychometric curve according to the (signed) distance of the first tone from the mean ($f_1 - \langle f \rangle$).

To assess contraction bias in neurotypical, dyslexia and ASD populations, we used the data of two experiments. Experiment 1 was administered to individuals with dyslexia and matched (age and reasoning) neurotypical (good reading) participants. Its results were previously published[18] and are here re-analyzed. Experiment 2 was administered to participants with ASD (high functioning with no concomitant language difficulties), and to matched (age and reasoning) neurotypicals. The cognitive profile of these participants is reported in **Supplementary Table 1**. In both experiments, participants were asked to perform serial 2-tone frequency discrimination, though the specifics of the protocols slightly differed (detailed in Materials and Methods).

For each individual, we extracted a summary estimate of the magnitude of the contraction bias as the difference between the average shifts in the psychometric curves, calculated for trials where $f_1 > \langle f \rangle$ and trials where $f_1 < \langle f \rangle$ (average probit bias $\widehat{b}$; see Materials and Methods; illustrated in **Fig. 1c**). All groups had a significant contraction bias ($\hat{b}$ dyslexia: mean $= 0.24 \pm 0.02$ SEM, $t_{27} = 11.1,\ p < 10^{-12}$; $\hat{b}$ matched neurotypicals: mean $= 0.32 \pm 0.03$ SEM, $t_{29} = 11.44,\ p < 10^{-12}$; $\hat{b}$ ASD, mean $= 0.15 \pm 0.03$ SEM, $t_{36} = 4.32, p < 0.001$; $\hat{b}$ matched neurotypicals, mean $= 0.24 \pm 0.03$ SEM, $t_{31} = 8.65$, $p < 10^{-10}$; paired t-tests; **Fig. 1d**). Yet, in line with previous studies[11,19–21], the magnitude of the bias was significantly smaller in both the ASD and the dyslexia groups compared with their corresponding, matched neurotypical groups (two sample t-tests; $t_{56} = 2.07,\ p = 0.04$, Cohen's d $= 0.53$ for neurotypicals vs. individuals with dyslexia; $t_{67} = 2.05, p = 0.04$, Cohen's d = 0.49 for neurotypicals vs. individuals with ASD; **Fig. 1d**).

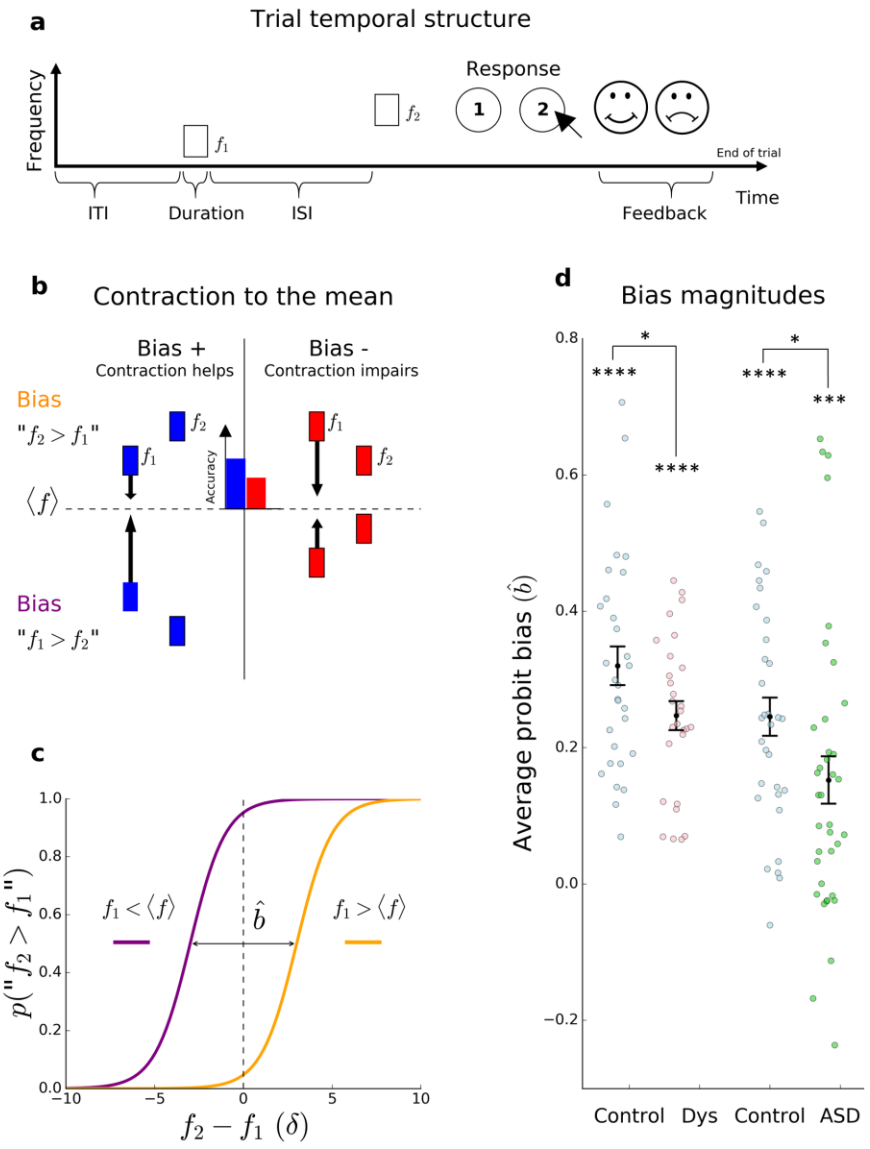


**Fig. 1. Contraction bias of individuals with dyslexia and individuals with ASD is smaller than neurotypicals'. (a)** Illustration of the temporal structure of a single trial. Two pure tones were presented sequentially, with a quiet inter-stimulus-interval (ISI). Participants then decided whether the 1st or 2nd tone had a higher pitch by pressing "1" or "2" and received a visual feedback. **(b)** Illustration of the contraction bias - contraction of the representation of the 1st tone towards the mean of the previous tones (dashed line). When the first tone ($f_1$) is closer to mean (left, blue trials), above (top) or below it (bottom), the perceived difference between the stimuli increases, and success rate increases (Bias+ trials). When the 2nd tone ($f_2$) is closer (right, red trials), this contraction reduces the perceived difference and accuracy decreases (Bias− trials). **(c)** Illustration of the probability to respond "$f_2 > f_1$" as a function of the difference between the tones $\delta$ (psychometric curve), plotted separately for the trials above ($f_1 > \langle f \rangle$; orange) and below ($f_1 < \langle f \rangle$; purple) the mean (b). Contraction bias can be

measured as the shift between the curves (average probit bias $\hat{b}$; see Materials and Methods). **(d)** Contraction bias in the dyslexia (red; $n = 28$) compared with matched neurotypical (blue; $n = 30$), and in ASD (green; $n = 37$), compared with matched neurotypical (blue; $n = 32$) groups. Data points denote individual participants. Though all participants of Experiment 1 (left plots), and most participants of Experiment 2 (right plots) showed contraction bias, it was significantly smaller in the dyslexia ($p = 0.04$) and ASD groups ($p = 0.04$). Black points indicate the mean values across subjects; error bars indicate standard error of the mean. $* \; p < 0.05$; $*** \; p < 0.001$; $**** \; p < 10^{-4}$; two sample t-tests (two-sided).

**Perceptual bias is dominated by recent stimuli in dyslexia and by longer-term statistics in ASD.**

In the analysis above contraction bias was calculated with respect to the overall mean of past stimuli. However, previous studies have shown that behavior of neurotypical participants depends on both recent stimuli (a phenomenon often termed serial dependence[17,22,23]), and the overall stimulus distribution[16,24]. Hence the shaping of perception by past stimuli may be better modelled as a combination of these contributions[17]. This combination may reflect an assumption about the stability of the environment. When the environment is stationary, the equally weighted empirical distribution built out of past samples, provides the best estimate of the current stimulus statistics. However, in a volatile environment, more recent stimuli are more representative of current statistics and so should be given larger weights. Therefore, the relative contributions of longer-term versus recent statistics reflect the balance between robustness and adaptability.

To estimate the relative contributions of recent versus longer-term statistics to participants' performance, we fitted a Generalized Linear Model (GLM, see Materials and Methods) to our data. The model included longer-term (distance of current 1st tone from the mean frequency

of all trials; $d_{\infty}^{t}$ in **Fig. 2a**) and recent contributions (distance of current 1st tone from the mean frequency of recent trials, e.g. $d_{1}^{t}$, $d_{2}^{t}$ for 1 and 2 trials back, **Fig. 2a**) as separate terms (shared for all participants). In both neurotypical groups, the distance from the previous trial's frequency and the distance from the overall mean frequency were the two main contributors to the bias, in line with previous results[17,25]. This can be seen in the larger magnitudes of the estimated weights ($w_1$ and $w_{\infty}$) associated with the trial-specific history terms $d_{1}^{t}, d_{\infty}^{t}$ (**Fig. 2b-c**; Materials and Methods; the ratio between $w_1$ and $w_{\infty}$ depends on the specific stimuli distribution, as explained in the next section. Hence, each population is compared to its matched neurotypical group, who performed exactly the same protocol). Both the dyslexia group and the ASD group differed from their corresponding neurotypical groups, but in different ways. The bias of the dyslexia group towards the frequencies of the most recent trial (hereon termed bias-by-recent) was similar to that of the neurotypical group, but their bias towards longer-term statistics (hereon termed bias-by-longer-term) was smaller than that of the neurotypical group ($w_{\infty}$ dyslexia: mean $= 0.51 \pm 0.09$ SEM; neurotypical: mean $= 0.89 \pm 0.1$ SEM; $p = 0.007$, permutation test; **Fig. 2b**). The ASD group showed the opposite pattern. The bias-by-recent of participants with ASD was smaller than that of neurotypical participants ($w_1$ ASD: mean $= 0.4 \pm 0.12$ SEM; neurotypical: mean $= 0.93 \pm 0.14$ SEM; $p = 0.04$, permutation test; **Fig. 2c**), but their bias-by-longer-term did not differ from neurotypicals'.

To assess individual variability, for each participant we introduced to the fit a random intercept and a random weight for each of the two main covariates $d_1$ and $d_{\infty}$ (random effects, Materials and Methods) in addition to the shared (group) parameters, and compared the individual weights of $d_1$ and $d_{\infty}$ in the neurotypical and test groups. In Experiment 1, the only weight that differed between individuals with dyslexia and their matched neurotypicals

was the contribution of longer-term statistics (Mann-Whitney $U = 559, p = 0.03$, Cohen's d = 0.61; **Fig. 2d**). In Experiment 2, the only weight that significantly differed between the ASD group and their matched neurotypicals was the contribution of recent stimuli (Mann-Whitney $U = 965, p < 10^{-6}$, Cohen's d = 1.22; **Fig. 2e**).

These two inter-group comparisons revealed a double dissociation in the dynamics of acquisition of environmental statistics. The magnitude of participants with dyslexia's bias-by-recent was similar to that of their matched neurotypicals, but their bias-by-longer-term had a smaller contribution. By contrast, participants with ASD were less affected by recent events, but showed the same magnitude of longer-term contribution as their matched neurotypical participants.

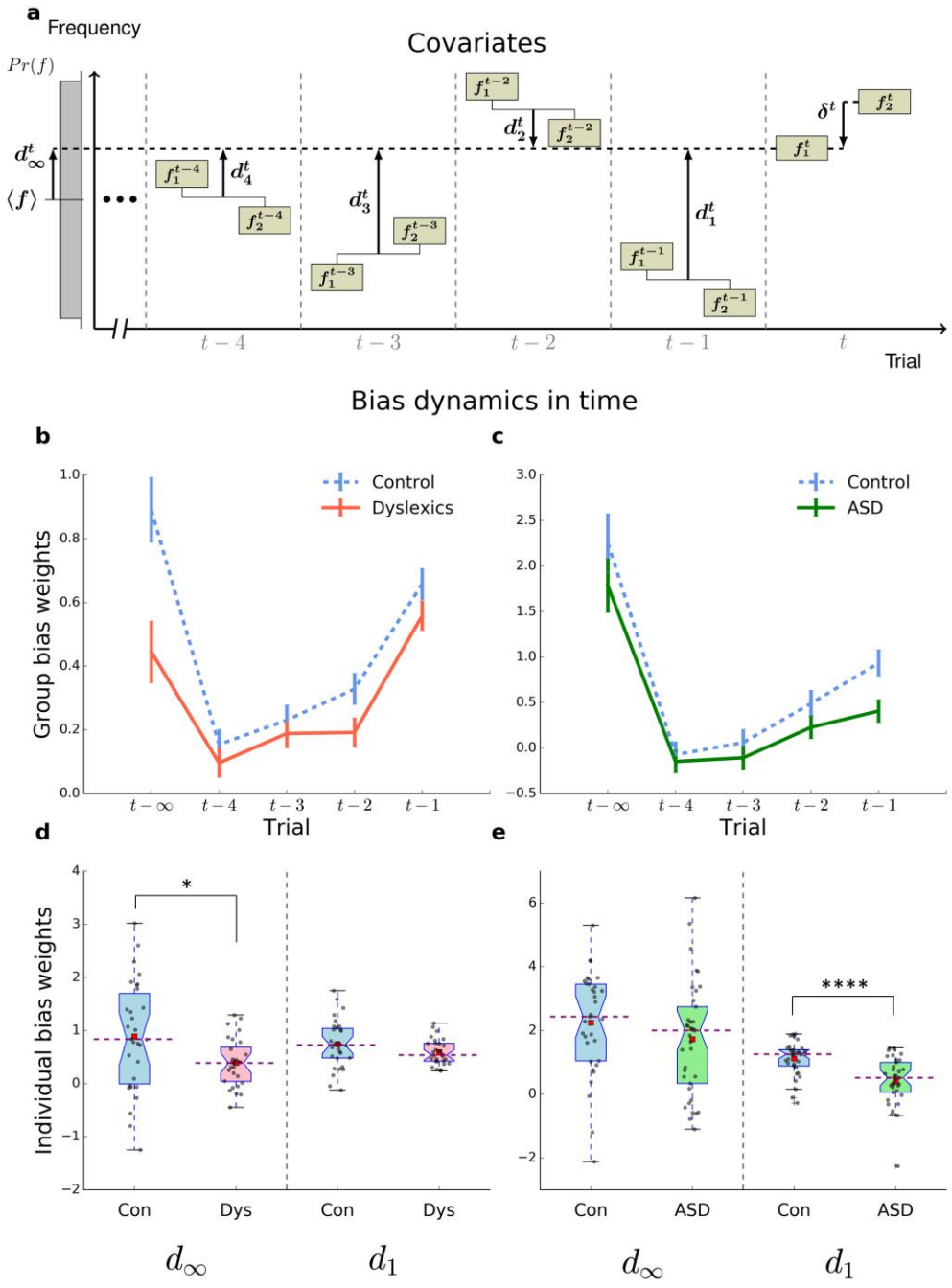


**Fig. 2. The dyslexia and ASD groups show opposite dynamics of contributions by recent and by longer-term statistics to their contraction bias**. **(a)** Illustration of the covariates of the GLM model. Covariates capturing recent sensory history in trial $t$ are defined as the frequency distance between the frequency of the first tone in trial $t$ and the average of the two tone frequencies at each time window; e.g. $d_1$ is the distance to the previous trial's mean frequency; $d_\infty$ is a special covariate, defined as the centered absolute frequency difference of the first tone in trial $t$. **(b-c)** The contribution of each covariate in each group. Each data point shows the estimate (mean) calculated for the

aggregated data of all participants. **(b)** The estimated group weights for participants with dyslexia (red; $n = 28$) and neurotypicals (blue; $n = 30$). Weights of the four most recent trials ($w_1, \ldots, w_4$) in the dyslexia group are equal to neurotypicals' but the weight of the bias-by-longer-term ($w_\infty$) is smaller. **(c)** The estimated group weights for ASD (green; $n = 37$) and neurotypical (blue; $n = 32$) groups. The weight of the bias-by-recent ($w_1$) among participants with ASD is smaller than neurotypicals' whereas their other weights are similar. Error bars indicate standard deviation. **(d-e)** group differences between $w_1$ and $w_\infty$ weights fitted individually. **(d)** The individually estimated weights of bias-by-recent (right) and bias-by-longer-term (left), of participants with dyslexia (red; $n = 28$) and neurotypicals (blue; $n = 30$) in Experiment 1. The bias-by-longer-term of individuals with dyslexia is smaller than neurotypicals' ($p = 0.03$, two-sided Mann-Whitney). **(e)** Same estimates for the participants with ASD (green; $n = 37$ and neurotypicals (blue; $n = 32$) of Experiment 2. Here only the bias-by-recent significantly differs between the groups, with larger estimates for neurotypicals ($p < 10^{-6}$, two-sided Mann-Whitney). Red squares and red horizontal lines indicate the mean and median values of each group, respectively. Error bars show lower to upper quartile values of the data. $* \; p < 0.05, \; **** \; p < 10^{-4}$.

## Neurotypical bias functions

The GLM assumes that each contribution to the bias depends linearly on the corresponding frequency difference. This assumption enables an approximated estimation of the relative contribution of current, recent and earlier stimuli to perception. However, it does not enable a characterization of the complexity with which history is internalized. Such characterization is crucial for understanding how on-going bias relates to the gradual internalization of task-relevant statistics that underlies expert level performance. Indeed, there are theoretical reasons to expect non-linear longer-term dependence[16]. Yet, previous studies have only used linear methods[11,17,21,23,26]. To study the bias with non-linear modeling, we conducted a new experiment online via the Mechanical Turk platform (Materials and Methods; **Supplementary Table 2**), and administered four different stimuli distributions, to four large groups of neurotypical participants, respectively. After replicating all the basic contraction

bias results (**Supplementary Fig. S1**), we used a Generalized Additive Model (GAM) with probit link function to analyze these data. The GAM replaces each linear term $w_i d_i^t$ (where $i = 1,2,3,4$ or $\infty$) of the GLM with an arbitrary function $b_i(d_i^t)$ (estimated using thin-plate regression splines[27]). Due to the increase in estimation complexity, attaining a reliable GAM analysis requires reasonably consistent and less noisy performers. This was obtained by including only participants with 65% accuracy and above (excluding ~15% of the participants).

We found the following non-linear frequency dependencies of the bias. The functional form of bias-by-recent ($b_1$, **Fig. 3a-d left**) was similar in all four distributions (see also serial biases in visual perception[22,23,28]). It increased gradually with distance from the current (first) tone, peaked at 0.4-0.5 octaves away, and then gradually decreased. A similar shape characterized the bias towards the frequencies of 2,3,4, and 5 trials earlier ($b_2 - b_5$), but its magnitude quickly decreased across trials (**Supplementary Fig. S2**). Still, the remaining trials had a significant contribution to the bias' variance ($b_\infty$, i.e. all other trials, longer-term effect), even when $b_1 - b_5$ were calculated separately.

The bias-by-longer-term was distribution sensitive, as evident in the distribution-specific patterns (**Fig. 3a-d right**, and verified by cross-validation; **Supplementary Fig. S3**). These patterns can be intuitively understood when stimuli distribution is considered. For uniform distributions (**Fig. 3a-b**), given a noisy measurement $f$ in the middle of the distribution, all plausible true stimuli around this measurement are a-priori equally likely, hence there should be no bias in this region. However, for noisy measurements at the edges of this distribution, the prior is crucially informative, since it excludes values 'outside' the prior, pulling perception towards plausible, true stimulus values (consistent with an observed contraction

bias[16]). When the uniform distribution is narrower (**Fig. 3b** – 2 octaves wide versus **Fig. 3a** - 3 octaves wide), the position of the edges is shifted (half an octave in each direction), but the qualitative shape of the bias function remains unchanged. With a Gaussian distribution, we expect a linear bias, pulling towards more probable values (**Fig. 3c**). With two (bimodal) uniform distributions we expect two typical uniform curves – one for each mode (**Fig. 3d**).

Using the same estimation method for our participants and for an Ideal (Bayesian) Observer (IO)[16], we could also quantitatively compare the contribution of the bias functions. These shapes are indeed optimal and characterize the shapes of the IO, whose sensory sensitivity ($\alpha$) is that of our human performers, and has full knowledge of stimuli distribution (Materials and Methods; **Supplementary Fig. S4**). **Fig. 3a-d right** show that the human (blue) and the IO (superimposed in black lines) bias functions roughly overlap in each of the four tested distributions. Note that the enhanced bias-by-recent (and one-two earlier ones) compared with all other trials is advantageous when the statistics changes, but is not optimal in our protocols, with stationary statistics. Hence, we don't calculate the IO model of the bias-by-recent.

Before applying the GAM analyses to our test populations, we verified the following. First, we validated the advantage of the GAM functions compared with the GLM weights. Applying cross-validation, we found that for each distribution, GAM significantly increased the predictive power of the fits ($w < 2, p < 0.01$ in all four distributions, Wilcoxon Signed-rank, cross-validation; **Supplementary Fig. S5**). Second, we verified that contraction bias is reliably more predictive than response inertia bias (the tendency to repeat the same response after a success and switch it after a failure[29,30]; $w < 1, p < 0.01$ in all four distributions, Wilcoxon Signed-rank, cross-validation; **Supplementary Fig. S5**). Third, we verified that

assuming interactions between bias-by-recent and by bias-by-longer-term has no predictive advantage over additive summation of these functions (**Supplementary Fig. S5**).

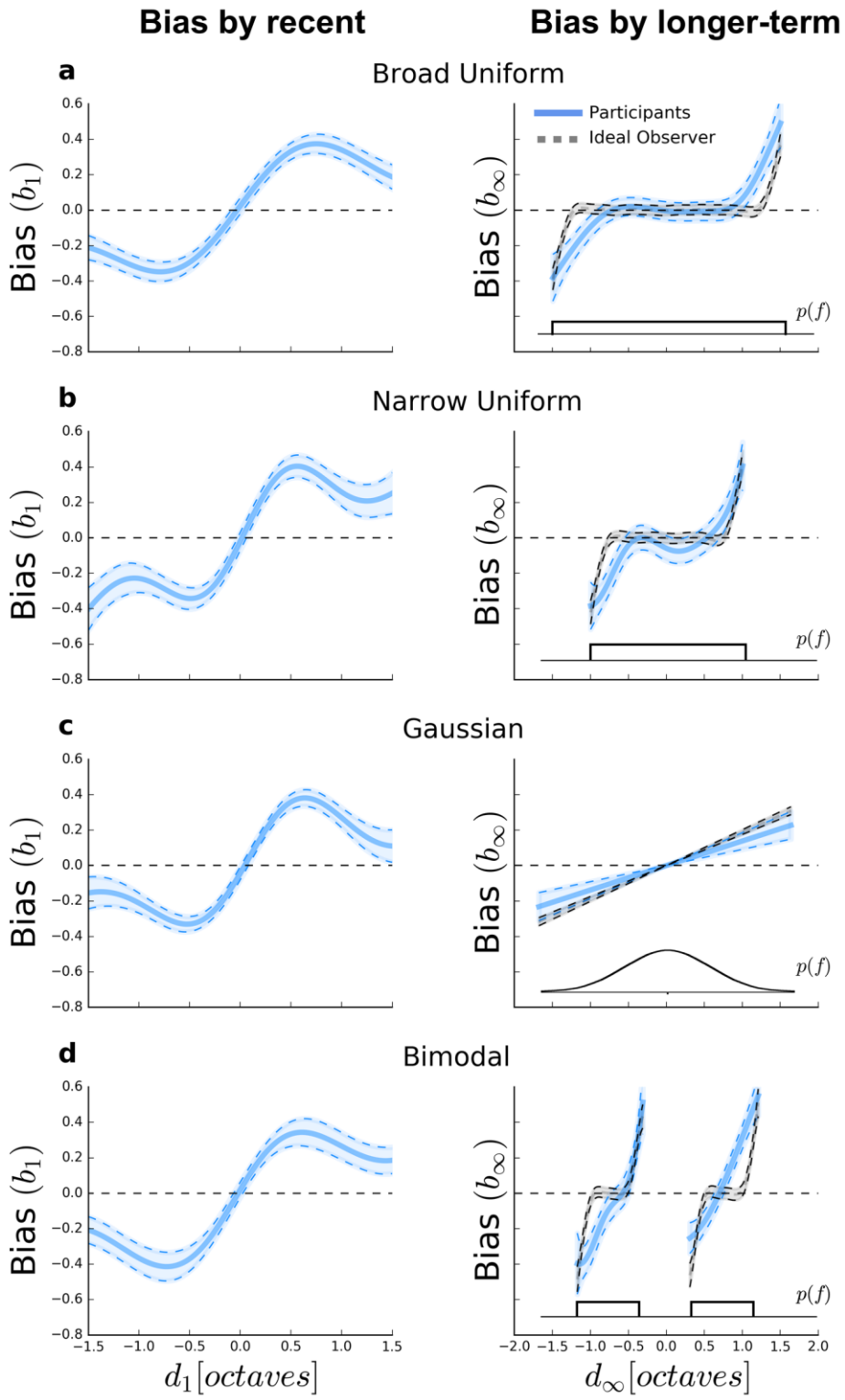


**Fig. 3. Bias-by-recent and bias-by-longer-term stimuli statistics, estimated using Generalized Additive Model (GAM) for 4 different frequency distributions**. **(a)** Broad, 3 octave, uniform ($n =$

125), **(b)** Narrow, 2 octave, uniform ($n = 94$), **(c)** Gaussian ($n = 163$) and **(d)** Bimodal of two uniform (0.9 octaves each; $n = 108$). **Left –** Bias-by-recent functions ($b_1(d_1)$) estimated with GAM regression. Each curve describes the magnitude of the bias as a function of the frequency distance (in octaves) of the first tone in the current trial from the previous trial mean ($d_1$). Though estimated separately, the 4 functions are similar in shape and magnitude. At the extreme edges the correlation between the two covariates ($d_1$ and $d_\infty$) is very high, and thus estimated values are less reliable (most notably for "narrow uniform"). **Right –** Bias-by-longer-term ($b_\infty$ ($d_\infty$)) of each distribution. The x-axis shows the frequency distance (in octaves) of the first tone in the current trial from the mean frequency of the distribution ($d_\infty$). Black superimposed curves show IO bias, estimated with the same model for simulated responses with matched noise level (yielding ~85% correct responses) and with priors corresponding to the distribution of each experiment (Materials and Methods). In the plots on the right, the x-axis zero denotes the mean of the distribution and the other values denote the relative position of $f_1$ within the distribution, illustrated at the bottom of each plot (note that in the plots on the left, the x-values correspond to those of $d_1$, which differ from the relative position of $f_1$). In all plots error bars indicate 95% confidence intervals. Dashed line denotes zero bias.

**Bias-by-longer-term is optimal in ASD but sub-optimal in dyslexia**

With nonlinear GAM analyses (which yielded better predictions than GLM across data sets; **Supplementary Fig. S6**) we could now ask if indeed the bias-by-longer-term, reflecting an integration of the details of the stimulus distribution, is atypical and sub-optimal in dyslexia but preserved in ASD, and whether the bias-by-recent function of the ASD group deviates from that in the neurotypical group. We first administered the novel GAM analysis to the performance of participants with dyslexia and neurotypicals in Experiment 1. Stimuli distribution was Gaussian, and based on the large-scale M-Turk experiment with neurotypical participants, we predicted neurotypicals' linear (**Fig. 3c**) IO's bias-by-longer-term and typical bias-by-recent function. This was indeed the case for bias in neurotypicals (**Fig. 4a**). As predicted, dyslexia bias-by-recent was roughly identical to neurotypical (**Fig. 4b**-left). Yet, -

their bias-by-longer-term was sub-optimal (**Fig. 4b**-right). It was both small in magnitude and did not display a linear shape, suggesting that, at the edges of the distribution, the contraction of individuals with dyslexia is expected to be reduced.

To obtain a quantitative comparison for these non-linear estimates, we quantified the contribution of the bias by calculating its variance across trials, namely, the extent to which variability in predicted responses can be attributed to bias towards previous trials[31] ($Var_{(b_i)} = \langle (b_i^t - \langle b_i \rangle)^2 \rangle$, where $b_i^t = b_i(d_i^t)$ is the bias in trial $t$ and $\langle b_i \rangle$ is the average bias across all trials). The variance of the bias-by-recent did not differ between individuals with dyslexia and neurotypicals ($p = 0.37$, permutation test). To account for individual variability in both bias-by-recent and bias-by-longer-term, we again introduced random effects. For each participant, we fitted a random intercept (to account for a global response preference), and a random slope (which could tune the magnitude of the individual bias function) for each function. These 3 additions (individual random intercept and slopes) to the model significantly increased its predictive power (cross-validation; **Supplementary Fig. S7**). To assess the individual variance of the bias-by-longer-term ($b_\infty$), we compared each individual to its noise-matched IO. Indeed, the variance of the bias-by-recent for individuals with dyslexia did not differ from neurotypicals' ($U = 393, p = 0.31$, Mann-Whitney; **Fig. 4c**), but variances of the bias-by-longer-term did ($W = 3,\ p < 10^{-5}$, Cohen's d = 1.15, Wilcoxon Signed-rank; **Fig. 4d**). We also tested whether the difference between the contribution of the bias-by-recent and the (distance from) optimality of the longer-term component significantly differs between the two populations. We calculated the individual difference between the variance of the bias by-recent ($var_{(b_1)}$) and the distance between the empirical and IO bias-by-longer-term variances ($\Delta var_{(b_\infty)} = var_{(b_\infty)} - var_{(b_\infty^{IO})}$). This difference ($var_{(b_1)} -$

$\Delta var_{(b_\infty)}$), was significantly larger in the dyslexia group compared to the neurotypical group ($U = 184, p = 0.005$, Mann-Whitney, Cohen's d = 0.98).

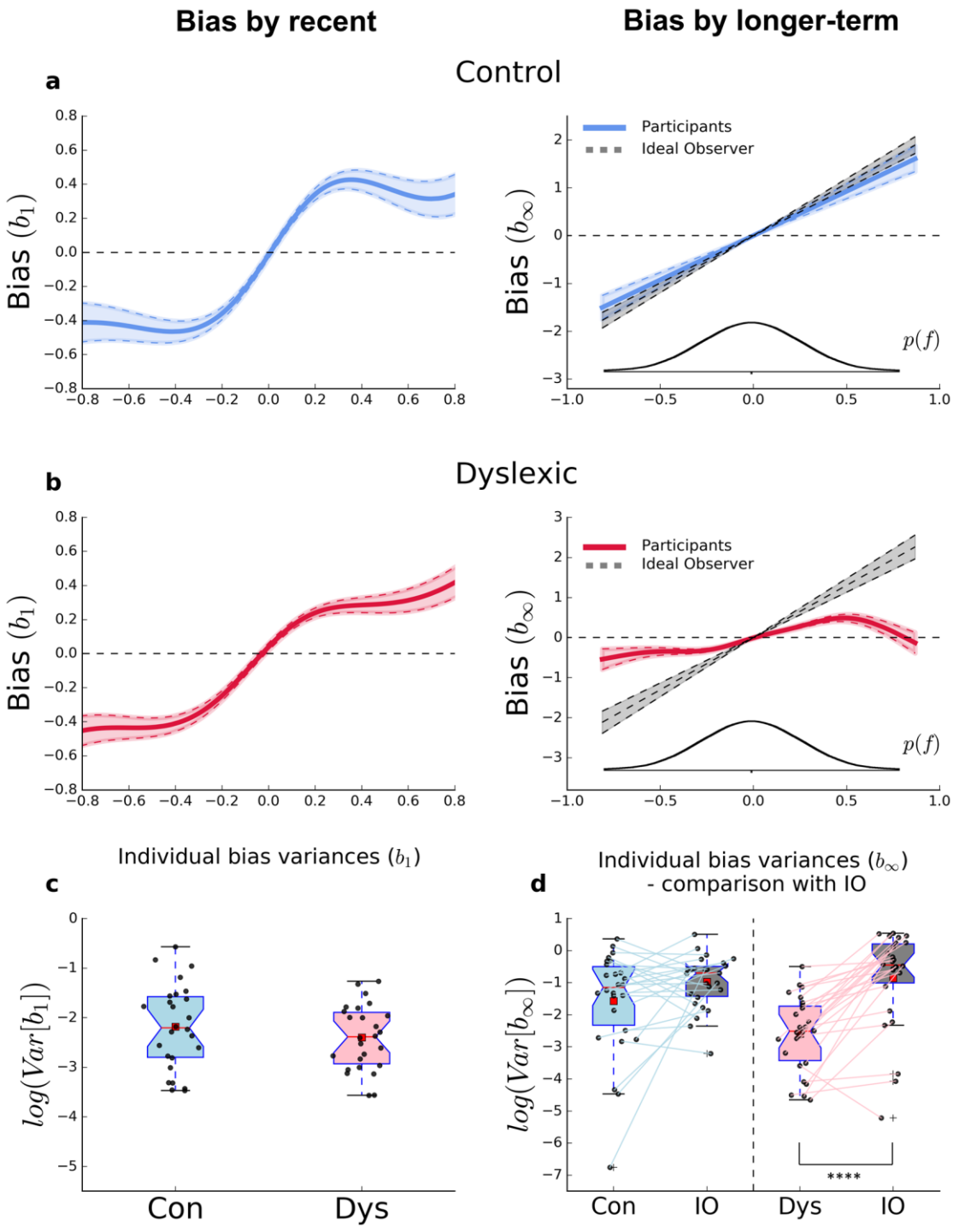


**Fig. 4. Normal bias-by-recent but abnormal bias-by-longer-term in individuals with dyslexia.** **(a-b) Left –** functions of bias-by-recent ($b_1(d_1)$) estimated for the (matched) neurotypicals **(a)** (blue; $n = 26$) and participants with dyslexia **(b)** (red; $n = 25$) who performed Experiment 1. GAM exclusion was based on a 65% accuracy criterion; see Materials and Methods. Estimation at the edges is noisy

due to high correlations with longer-term distances. **Right –** functions of bias-by-longer-term ($b_\infty(d_\infty)$). The functions of noise-matched Ideal Observer (IO) are superimposed. In all plotted curves, error bars indicate standard deviation. **(c-d)** Individual bias variances, shown in logarithmic scale. **(c)** Variance of bias-by-recent across trials $Var(b_1)$. Bias-by-recent is similar in participants with dyslexia ($n = 25$) and neurotypical ($n = 26$) participants ($p = 0.31$, two-sided Mann-Whitney). **(d)** Variance of bias-by-longer-term $Var(b_\infty)$ of each group compared with their matched IO (black). Neurotypicals' bias ($n = 26$) is similar to their matched IO ($p = 0.38$, two-sided Mann-Whitney), whereas the bias of participants with dyslexia ($n = 25$) is smaller ($p < 10^{-5}$, two-sided Mann-Whitney). Red squares and red horizontal lines indicate the mean and median values of each group, respectively. Error bars show lower to upper quartile values of the data. $**** p < 10^{-4}$.

To calculate these bias functions for the ASD group we performed a new experiment (Experiment 3), since the frequency distribution in Experiment 2 was not sufficiently broad (0.65 octave). To allow a direct comparison between the bias functions in ASD and dyslexia, we recruited, in addition to individuals with ASD and neurotypicals, also individuals with dyslexia (**Supplementary Table S3** describes the cognitive profiles of the three, age and reasoning-matched groups) and administered the same protocol to these three groups. We chose a bimodal distribution for the common protocol (**Fig. 3d**) since it yields bias-by-longer-term within most of its equally sampled frequency range, and therefore reliable functions can be fitted with the smallest amount of trials. Importantly, its pattern of ideal contraction allows an evaluation of participants' resolution in integrating the distribution. Ideal contraction bias should be towards the mean of each mode rather than towards the global mean (mean across all trials). Contraction to the global and to the local means have the same direction at the edges, but opposing directions at the "inner" sections of each mode. Thus, adequate learning of the bimodal distribution is indicated by contraction to the local mean at the inner sections.

Neurotypicals (**Fig. 5a**) showed both the expected pattern of bias by-recent and the expected pattern of bias-by longer-term in bimodal distribution (**Fig. 3d**). Qualitatively, their bias-by-longer-term was directed towards the local means at the inner sections of the bimodal distribution. Quantitatively, variance comparison of this bias component showed neither group differences (black line superimposed, **Fig 5a**, right), nor individual differences (introducing the same random effects components as in Experiment 1; **Fig. 5e**, left) with respect to their noise-matched IOs. The group with ASD showed a significantly smaller bias-by-recent (**Fig. 5b**, left) compared with both the neurotypical group ($p = 0.03$, permutation test on group variance) and the group with dyslexia ($p$ = 0.03, permutation test on group variance), though its basic shape was similar. A comparison between the individual bias variances revealed a significant difference between the bias-by-recent of the 3 groups ($H = 37.8, p < 10^{-9}$ Kruskal-Wallis; **Fig. 5d**), due to a smaller variance in the ASD group (a follow-up Dunn's multiple comparisons test, $Z > 4.3, p < 10^{-5}$, bonferroni correction, Cohen's d > 1 for the two comparisons). However, their bias-by-longer-term was adequate, indicating detailed integration of stimuli distribution, and quantitatively overlapped that of matched IOs (**Fig. 5b**, right). Dyslexia bias-by-recent was similar to neurotypical (**Fig. 5c**). In fact, the variance of their estimated bias was even slightly larger than neurotypicals' ($z = 2.2, p = 0.04$, bonferroni correction, Dunn's test). Cross-validation on the aggregated data of the three populations showed that estimating separate bias-by-recent functions for neurotypicals and individuals with dyslexia does not improve the predictive power of the model ($W = 24,\ p = 0.72$, Wilcoxon Signed-rank, cross-validation), though estimating a separate bias-by-recent function in the ASD group did improve it ($W < 1,\ p = 0.009$, Wilcoxon Signed-rank, cross-validation).

Yet, bias-by-longer-term in the dyslexia group was weaker, and deviated from the matched IOs. Qualitatively, in the "inner" parts of the two modes, the bias of participants with dyslexia was much smaller than the expected IO's bias, indicating sub-optimal use of previous history. Quantitatively, while the individual biases of neurotypicals and of participants with ASD were similar to their matched IO's, in dyslexia, they were significantly smaller ($W = 9,\ p = 0.006$, Cohen's d = 0.85, Wilcoxon Signed-rank; **Fig. 5e**). Additionally, the difference between the contribution of the bias-by-recent and the (distance from) optimality of the longer-term component significantly differed between the groups ($H = 28.4, p < 10^{-7}$ Kruskal-Wallis). Post-hoc tests showed that this difference was larger in individuals with dyslexia than in neurotypicals ($z = 3.1, p = 0.002$, bonferroni correction, Dunn's test), and was smaller in individuals with ASD compared to neurotypicals ($z = 2.7, p < 10^{-7}$, bonferroni correction, Dunn's test). Performance of individuals with dyslexia in this experiment was mildly poorer than that of neurotypicals and individuals with ASD (**Supplementary Table S3**). Hence, the Bayesian framework predicts a larger contraction bias for the dyslexia group, reflecting increased use of previous statistics . However, their actual bias was smaller than expected (quantified by their noise-matched IO) given their reduced frequency sensitivity.

To summarize, the sub-optimal bias functions of participants with dyslexia suggest that they do not adequately integrate information from longer-term statistics. By contrast, the bias-by-longer-term of participants with ASD is optimal, suggesting adequate ability to learn complex distributions. Yet, their ability to quickly adapt to changes is reduced as reflected in the smaller magnitude of their bias-by-recent function.

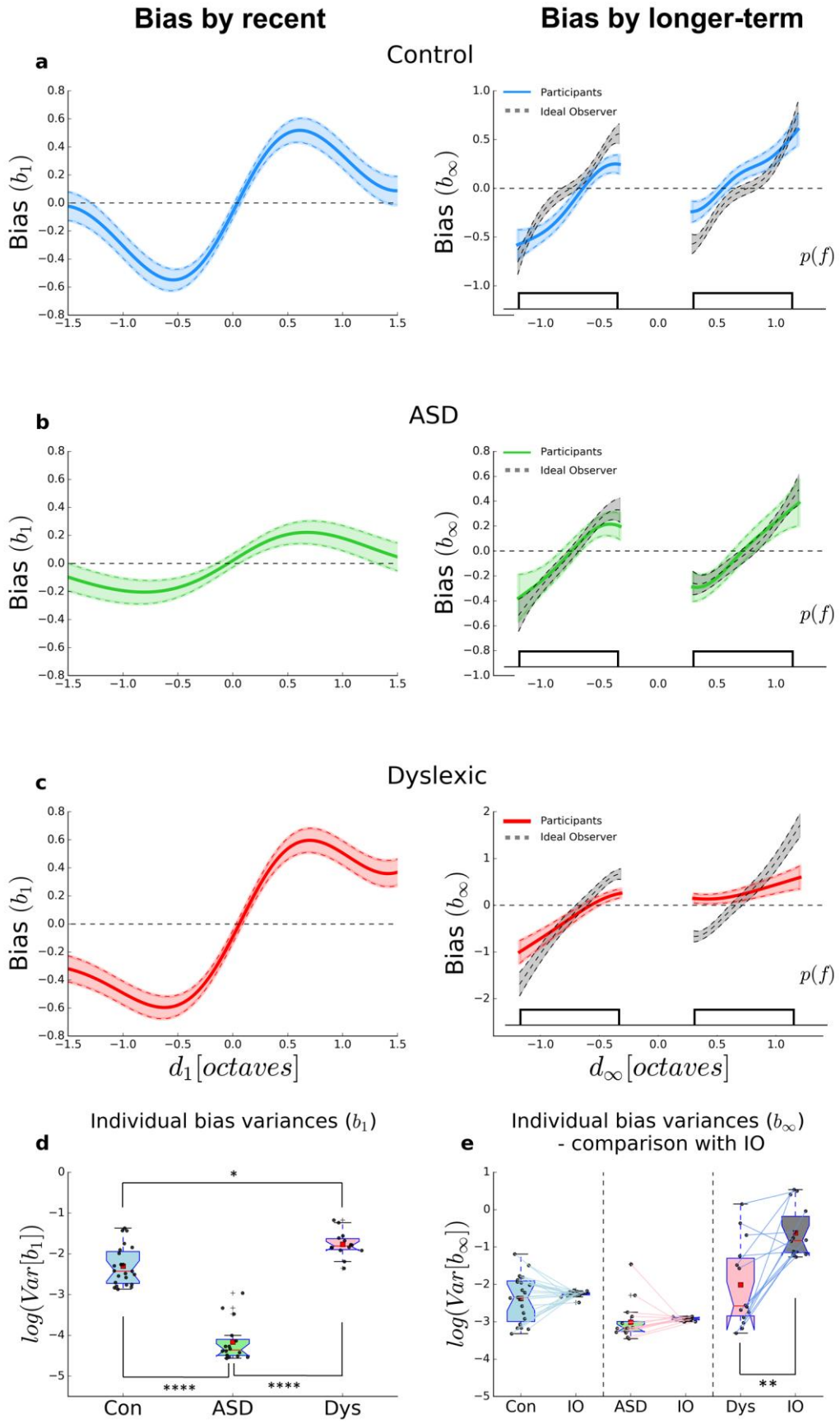
Bias by recent
Bias by longer-term
a
Control
Bias ($b_1$)
Bias ($b_\infty$)
Participants
Ideal Observer
$p(f)$
b
ASD
c
Dyslexic
$d_1[octaves]$
$d_\infty[octaves]$
Individual bias variances ($b_1$)
Individual bias variances ($b_\infty$)
- comparison with IO
d
e
$log(Var[b_1])$
$log(Var[b_\infty])$
Con
ASD
Dys
IO
*
****
**

**Fig. 5. Functions of bias-by-recent and bias-by-longer-term estimated for neurotypical, ASD and dyslexia groups performing the same experimental protocol.** **(a)** neurotypicals ($n = 23$), **(b)** participants with ASD ($n = 16$) and **(c)** participants with dyslexia ($n = 14$) who participated in Experiment 3 (bimodal distribution). **Left –** The magnitude of the bias-by-recent as a function of the frequency distance ($b_1(d_1)$) of the first tone in the current trial from the previous trial mean ($d_1$). **Right –** Bias-by-longer-term ($b_\infty(d_\infty)$). The x-axis denotes the frequency distance of the first tone in the current trial from the mean frequency of the distribution in octaves ($d_\infty$). Black superimposed curves show Ideal Observer (IO) bias. The pattern of bias in both neurotypicals **(a)** and participants with ASD **(b)** is roughly overlapping the IO's. The pattern of bias-by-longer-term for the dyslexia group **(c)** deviates from their IO, with an overall smaller magnitude. Error bars indicate standard deviation. **(d-e)** Individual bias variances of the groups that participated in Experiment 3, shown in log scale. **(d)** Variance across trials of bias-by-recent $Var(b_1)$ is a bit larger for participants with dyslexia ($n = 14$) compared to neurotypicals ($n = 23$; $p = 0.04$, Dunn's multiple comparisons test with bonferroni correction), but substantially smaller in ASD ($n = 16$) compared to neurotypicals ($p < 10^{-9}$, Dunn's multiple comparisons test with bonferroni correction). **(e)** Variances of bias-by-longer-term ($Var(b_\infty)$) of each group compared with their noise-matched IOs. The bias of participants with dyslexia ($n = 14$) is smaller than that of their matched IOs ($p = 0.006$, two-sided Mann-Whitney), whereas the bias of neurotypicals ($n = 23$) and participants with ASD ($n = 16$) is similar to their matched IOs' ($p = 0.8$, and $p = 0.09$ respectively, two-sided Mann-Whitney). Red squares and red horizontal lines indicate the mean and median value of each group, respectively. Error bars show lower to upper quartile values of the data. $* \; p < 0.05$, $** \; p < 0.01$, $**** \; p < 10^{-4}$.

## Discussion

Perceptual decisions of both individuals with dyslexia and individuals with ASD were less affected by previous trials compared to neurotypicals, in line with previous studies (refs 18,19 for dyslexia, and 10,11,21 for ASD). However, individuals with dyslexia are "fast updaters" and "fast forgetters" compared to both neurotypicals and individuals with ASD.

Individuals with ASD exhibit the reverse pattern. They are “slow updaters” and “slow forgetters” compared to neurotypicals (and individuals with dyslexia). Using Generalized Additive Model (GAM) enabled us to assess the detailed non-linear bias functions of these two temporal scales and to characterize the integration of the statistics underlying long-term representations. These analyses revealed that the bias towards recent stimuli does not depend on the specific stimuli distribution and is largely confined to neighboring (~0.5 an octave away) frequencies in all three groups of participants. Yet this bias is smaller in ASD, limiting the rate of adaptation to new statistics. The bias-by-longer-term observed in neurotypicals (more than a few seconds) reflects optimal use of the information conveyed by the distribution, well beyond the mean and standard deviation, and largely overlapping that of an IO. The bimodal distribution of pure tones in Experiment 3 can be thought of as a toy example of two distinct categories, which individuals with dyslexia failed to internalize accurately (as suggested by the sub-optimal bias they manifested) in contrast to neurotypicals and individuals with ASD. We propose that the resolution of the learned distribution is indicative of the quality with which categorical representations (e.g. of phonemes/syllables) form, at least within session. For all three groups of participants, future research should study the accumulation versus forgetting of stimuli information across days and years. Our GAM analyses provide the tools, and the comparison to IO quantifies optimality.

Remarkably, the patterns of perceptual biases parallel the unique pattern of difficulties and strengths of each of these two populations. Individuals with dyslexia are diagnosed based on a difficulty in efficient decoding of the written text, which relies on long-term regularities of spoken and written language. Failing to efficiently benefit from these regularities may pose a main impediment to their ability to form rich categorical representations (reviewed in[7]) and

expert-level reading skills[3]. The resolution of our categorical representations is not typically challenged in daily situations, which contain many cues. But it can be tested in challenging ones, for example, when individuals are asked to identify the (not very familiar) speaker. Identification in one's native language, where categorical representations of speech sounds are rich and robust due to extensive exposure, is more accurate than in a second language, where speech sounds are less familiar. Individuals with dyslexia are known to have difficulties in second language acquisition. Yet, counter-intuitively, their performance in the challenging task of identifying the speaker's identity is particularly impaired (compared with neurotypicals) in their native language: the familiarity advantage afforded by one's native language is smaller in dyslexia compared with neurotypicals, suggesting slower exposure-based enrichment of categorical representations (at least for speech)[6]. We predict that individuals with dyslexia will manifest a similarly reduced advantage of familiarity also in other, non-linguistic domains.

By contrast, the benefit from repeating recent stimuli (<3-4 sec) is similar in neurotypicals and individuals with dyslexia for various types of stimuli, including written words and non-words[18]. However, when the time intervals between consecutive encounters of the same non-word increase, this benefit decays faster in individuals with dyslexia than in neurotypicals[18].

Individuals with ASD are less affected by recent information[21] and show slower adaptation[10,32] compared to neurotypical individuals. They show a dissociation between adequate learning of longer-term statistics, as suggested by their intact categorical representations[14], and impaired integration of recent information. Their slow updating of recent information is manifested in both social[33] and cognitive contexts. For example, Happé[34] characterized oral sentence-reading using words that have more than one correct

pronunciation: one that is generally (long-term) more common, and one that is context-wise (temporally local) more appropriate. Individuals with ASD showed a greater preference for the generally more common pronunciation. A conceptually similar observation was found in our participants with ASD (**Supplementary Table S1**), who did not differ significantly from neurotypicals and were faster than individuals with dyslexia in decoding of pseudo words, which have no recent context. However, they were slow in paragraph reading, which should benefit from the semantic context of the recently read words. We propose that individuals with dyslexia rely on recent context to compensate for their slow word decoding, whereas individuals with ASD partially compensate for their slow use of the semantic context by using long-term regularities.

As explained above, the "slow update" of internal representations in ASD is in line with the central coherence hypothesis[34,35], and explains the difficulty in adequate context-specific responses. Yet, it seems inconsistent with the hypothesis that individuals with ASD overestimate the environment's volatility[36]. Hyper-volatility estimation is expected to yield larger weights to recent events, which are attributed greater information, whereas we found reduced weights. Similarly, our observations seem inconsistent with the hypothesis that individuals with ASD overweight their prediction error, which leads to overfitting their predictions, yielding reduced generalization[37]. We do not propose that individuals with ASD overfit. We find that they learn the stationary statistics, which are reliable across situations but might be less important when recent context is more relevant. Other studies also report that long-term priors are reliably learned by individuals with ASD[38], and even several seconds' intervals seem to suffice for their learning of partially novel stimuli[39]. We suggest that individuals with ASD are 'slow adapters'. We predict that they will be slower in error correction and in tracking changes in the external environment, as will be manifested in both

perceptual and sensorimotor tasks. Still, understanding the differences between the implications of the different studies probably requires assessment of bias under several time windows of changes in environment's statistics (better characterizing the term volatility), and under several levels of reliability of the statistics of each environment at any given moment.

Importantly, the 'slow-update' proposal explains seemingly unrelated observations in the visual modality. It has been proposed that individuals with ASD have a spatially local preference[40]. This conclusion was supported by poorer performance in global motion tasks in individuals with ASD. Yet, a recent meta-analysis[37] showed that this difficulty can be explained by slower spatial integration. Thus, the relative local advantage attributed to ASD may, in fact, reflect slow temporal updates of internal representations from spatially local to spatially more global stimuli[41].

In summary, the conceptual dissociation that emerges from the slow-updating and fast-forgetting characteristics of the perceptual biases observed in ASD and dyslexia, respectively, captures a broad range of seemingly unrelated phenomena within a unified framework. Importantly, it suggests a direct link between computational analyses and clinical profiles and applications. Specifically, it implies that a reduction in the rate of introducing new contexts might improve the effective integration and comprehension of these contexts by individuals with ASD, whereas enhancing regularities may facilitate the performance of individuals with dyslexia.

## Code Availability

We used the Psychtoolbox-3 MATLAB toolbox (http://psychtoolbox.org/) for creating and running experiments 2 and 3. The code for the online experiment which was applied via Mechanical-Turk is available at https://github.com/ItayLieder/Mech_turk_2afc. Analysis was conducted using the "mixed GAM computation vehicle with automated smoothness estimation" (mgcv) free package https://cran.r-project.org/web/packages/mgcv/index.html.

**Acknowledgments:** We thank Dr. Tamir Epstein, Tamar Malinovich, Gal Vishne, Odeya Guri, Shahaf Granot and Maayan Kurulkar for help collecting the experimental data.

**Funding:** This study was supported by the Israel Science Foundation (ISF grant no. 616/11 and Canada-Israel grant no. 2425/15) grant to Merav Ahissar, by the Gatsby Charitable Foundation, the German-Israeli Foundation for Scientific Research and Development (grant no. I-1303–105.4/2015) and grant to Itay Lieder (Young Researchers Exchange Scholarship) by The Jerusalem Brain Community (JBC).

**Author contributions:** I.L., initiated the project. I.L., S.J-D., designed the experiments. I.L., V.A., developed the model and analyzed the data used in this study. O.F., S.J-D., involved in data acquisition. M.A., M.S., conceptualization, supervision, investigation and methodology, funding acquisition. All authors contributed to the interpretation of data and writing of the paper and approved the final version of the manuscript for submission.

**Competing interests**

I declare that the authors have no competing financial or non-financial interests as defined by Nature Research

**References and Notes:**


1. Zoccolotti, P. *et al.* Word length effect in early reading and in developmental dyslexia. *Brain Lang.* **93,** 369–373 (2005).

2. Ahissar, M., Lubin, Y., Putter-Katz, H. & Banai, K. Dyslexia and the failure to form a perceptual anchor. *Nat. Neurosci.* **9,** 1558–1564 (2006).

3. Ahissar, M. Dyslexia and the anchoring-deficit hypothesis. *Trends Cogn. Sci.* **11,** 458–465 (2007).

4. Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T. & Kraus, N. Context-Dependent Encoding in the Human Auditory Brainstem Relates to Hearing Speech in Noise: Implications for Developmental Dyslexia. *Neuron* **64,** 311–319 (2009).

5. Oganian, Y. & Ahissar, M. Poor anchoring limits dyslexics' perceptual, memory, and reading skills. *Neuropsychologia* **50,** 1895–1905 (2012).

6. Perrachione, T. K., Tufo, S. N. Del & Gabrieli, J. D. E. Human Voice Recognition Depends on Language Ability. **333,** (2011).

7. Banai, K. & Ahissar, M. Poor sensitivity to sound statistics impairs the acquisition of speech categories in dyslexia. *Lang. Cogn. Neurosci.* **3798,** 1–12 (2018).

8. Pellicano, E. & Burr, D. When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16,** 504–510 (2012).

9. Sinha, P. *et al.* Autism as a disorder of prediction. *Proc. Natl. Acad. Sci.* **111,** 15220–15225 (2014).

10. Turi, M., Karaminis, T., Pellicano, E. & Burr, D. No rapid audiovisual recalibration in

adults on the autism spectrum. *Sci. Rep.* **6,** 2–8 (2016).

11. Karaminis, T. *et al.* Central tendency effects in time interval reproduction in autism. *Sci. Rep.* **6,** 1–13 (2016).

12. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub* (2013).

13. Newman, T. M. *et al.* Hyperlexia in children with autism spectrum disorders. *J. Autism Dev. Disord.* **37,** 760–774 (2007).

14. Bott, L., Brock, J., Brockdorff, N., Boucher, J. & Lamberts, K. Perceptual similarity in autism. *Q. J. Exp. Psychol.* **59,** 1237–1254 (2006).

15. Hollingworth, H. L. The Central Tendency of Judgment. *he J. Philos. , Psychol. Sci. Methods* **7,** 461–469 (1910).

16. Ashourian, P. & Loewenstein, Y. Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One* **6,** (2011).

17. Raviv, O., Ahissar, M. & Loewenstein, Y. How Recent History Affects Perception: The Normative Approach and Its Heuristic Approximation. *PLoS Comput. Biol.* **8,** (2012).

18. Jaffe-dax, S., Frenkel, O. & Ahissar, M. Dyslexics ' faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *Elife* 1–19 (2017). doi:10.7554/eLife.20557

19. Jaffe-Dax, S., Raviv, O., Jacoby, N., Loewenstein, Y. & Ahissar, M. A Computational Model of Implicit Memory Captures Dyslexics' Perceptual Deficits. *J. Neurosci.* **35,** 12116–12126 (2015).

20. Jaffe-dax, S., Lieder, I., Biron, T. & Ahissar, M. Dyslexics ' usage of visual priors is impaired. *J. Vis.* **16,** 1–9 (2016).

21. Molesworth, C., Chevallier, C., Happé, F. & Hampton, J. Children With Autism Do Not Show Sequence Effects With Auditory Stimuli. *J. Exp. Psychol. Gen.* **144,** 48–57 (2015).

22. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat. Neurosci.* **17,** 738–743 (2014).

23. Liberman, A., Fischer, J. & Whitney, D. Serial dependence in the perception of faces. *Curr. Biol.* **24,** 2569–2574 (2014).

24. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427,** 244–247 (2004).

25. Raviv, O., Lieder, I., Loewenstein, Y. & Ahissar, M. Contradictory Behavioral Biases Result from the Influence of Past Stimuli on Perception. *PLoS Comput. Biol.* **10,** 8–12 (2014).

26. Fassihi, A., Akrami, A., Esmaeili, V. & Diamond, M. E. Tactile perception and working memory in rats and humans. *Proc. Natl. Acad. Sci.* **111,** 2331–2336 (2014).

27. Wood, S. *Generalized Additive Models: an introduction with R*. (CRC press, 2006).

28. Kiyonaga, A., Scimeca, J. M., Bliss, D. P. & Whitney, D. Serial Dependence across Perception, Attention, and Memory. *Trends Cogn. Sci.* **21,** 493–497 (2017).

29. Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M. & Gardner, J. L. Adaptable history biases in human perceptual decisions. *Proc. Natl. Acad. Sci.* **113,** E3548–E3557 (2016).

30. Arzounian, D., De Kerangal, M. & De Cheveigné, A. Sequential dependencies in pitch judgments. *J. Acoust. Soc. Am.* **142,** 3047–3057 (2017).

31. Frund, I., Wichmann, F. A. & Macke, J. H. Quantifying the effect of intertrial dependence on perceptual decisions. *J. Vis.* **14,** 9–9 (2014).

32. Pellicano, E., Jeffery, L., Burr, D. & Rhodes, G. Abnormal Adaptive Face-Coding Mechanisms in Children with Autism Spectrum Disorder. *Curr. Biol.* **17,** 1508–1512 (2007).

33. McIntosh, D. N., Reichmann-Decker, A., Winkielman, P. & Wilbarger, J. L. When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Dev. Sci.* **9,** 295–302 (2006).

34. Happé, F. G. E. Central coherence and theory of mind in autism: Reading homographs in context. *Br. J. Dev. Psychol.* **15,** 1–12 (1997).

35. Frith, U. & Happe, F. The Weak Coherence Account : Detail-focused Cognitive Style in Autism Spectrum Disorders. *J. Autism Dev. Disord.* **36,** (2006).

36. Lawson, R. P., Mathys, C. & Rees, G. Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* **20,** 1293–1299 (2017).

37. Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit L., & Wagemans J. Precise Minds in Uncertain Worlds : Predictive Coding in Autism. *Psychol. Rev.* **121,** 649 (2014).

38. Croydon, A., Karaminis, T., Neil, L., Burr, D. & Pellicano, E. The light-from-above prior is intact in autistic children. *J. Exp. Child Psychol.* **161,** 113–125 (2017).

39. Van de Cruys, S., Vanmarcke, S. & Van de Put, I. The Use of Prior Knowledge for

Perceptual Inference Is Preserved in ASD. *Clin. Psychol. Sci.* **6,** (2018).

40. Mottron, L. & Burack, J. A. Enhanced perceptual functioning in the development of autism. In The development of autism: Perspectives from theory and research (pp. 131-148) (Mahwah, NJ, US, Lawrence Erlbaum Associates Publishers, 2001).

41. Robertson, C. E. & Baron-Cohen, S. Sensory perception in autism. *Nat. Rev. Neurosci.* **18,** 671–684 (2017).

## Materials and Methods

### Experimental design

#### Participants

*Participants with dyslexia and neurotypical participants - Experiments 1 and 3*

Performance and biases in the 2-tone frequency discrimination task were calculated by re-analyzing the data of adult individuals with dyslexia and neurotypicals matched in age and reasoning abilities, recruited and assessed by Jaffe-Dax et al. (Experiment 1[18]). Sixty native Hebrew speakers (30 individuals with dyslexia and 30 good readers), students at the Hebrew University, were recruited for Jaffe-Dax et al.'s experiment. All the participants with dyslexia had been diagnosed by authorized clinicians as having a specific reading disability. Reading-related measures were also assessed in our lab. For this analysis, we excluded two participants with dyslexia: one with accuracy in 2-tone discrimination < 50%; one with exceptionally low (negative) average probit bias (> 2.5 standard deviations). We report the data of the remaining 28 participants with dyslexia and 30 neurotypicals. Age, cognitive profile and discrimination performance of all (reported) participants of Experiment 1 are described in **Supplementary Table 1.** For Experiment 3 we invited individuals with dyslexia who participated in a previous experiment in the lab; 17 individuals returned to participate in this study, matched in age and cognitive skills (assessed by Block Design) to the two other

groups that participated in Experiment 3 (3 were excluded due to poor performance, <65% criterion chosen for GAM analysis explained below, based on our M-Turk experiment, retaining 14). The recruitment of the neurotypical participants for Experiment 3 is described in the following section. Age, cognitive profile and discrimination performance of all (reported) participants of Experiment 3 are described in **Supplementary Table S3**.

*Neurotypicals and participants with ASD – Experiments 2 and 3*

Recruitment of the participants with ASD was conducted through clinics, designated facilities and specific support groups. Recruitment of the neurotypical individuals was based on ads posted at the Hebrew University. All participants with ASD were clinically diagnosed with autism spectrum disorder according to DSM5 standards[12]. Forty one individuals with ASD and 35 neurotypicals were recruited for Experiment 2; Of the participants with ASD, two were excluded due to poor performance in the general reasoning task (scaled score of 4 and 3 in Block Design[42]), and 2 additional participants were excluded due to poor (chance level) performance in the 2-tone discrimination task. The data of the remaining 37 participants are reported. Of the neurotypical participants, two were excluded due to poor performance (50% accuracy) of the 2-tone discrimination task and one due to an exceptionally large average probit bias (> 2.5 standard deviations). We report the data of the remaining 32 participants. Age, cognitive profile and discrimination performance of all (reported) participants of Experiment 2 are described in **Supplementary Table S1**. 19 of the participants with ASD and 26 of neurotypicals also participated in Experiment 3. Three participants with ASD and 3 neurotypicals were excluded due to poor performance in 2-tone discrimination (<65% criterion, required for reliable GAM analysis, as calculated based on the M-Turk experiment), keeping 16 participants with ASD and 23 neurotypicals. Age, cognitive profile and discrimination performance of the (reported) participants of Experiment 3 are described in

**Supplementary Table S3**. In all three groups we only recruited participants with no or minimal (<2 years) formal musical experience, since previous studies have shown that individuals with musical experience perform better on frequency discrimination tasks[5]. Thus, all three groups of participants of Experiment 3, had no formal musical training, but participated in one previous 2-tone discrimination experiment in our lab.

All experiments were approved by both the Ethics committee of the Psychology Department of Hebrew University, and the Helsinki ethics committee of Sheba Hospital (required for testing for individuals with ASD recruited through their adult clinic). All participants provided written informed consent and were compensated financially for their time and travel expenses.

*Participants of the Mechanical-Turk (M-Turk) experiment*

Six hundred and sixty-four participants (all from the US) took part in four 2-tone frequency discrimination experiments (different stimulus distributions). All were recruited with previous M-Turk approval >95% and a total number of HITs (human intelligence task) > 1000. Our written instructions emphasized that the experiment must be performed: (1) using headphones in a quiet environment; (2) with either a laptop or a desktop computer and (3) only by people with good hearing who are between the ages of 20 to 50. Each individual could only participate once. They were given a payment of 2$ for 300 performed trials (corresponding to a duration of approximately 15-20 minutes). Participants gave their identifying M-Turk worker ID, age, gender and musical experience. We applied two performance-based inclusion criteria, based on: (1) mean performance and (2) performance consistency. Manipulating exclusion criteria, we found that excluding participants whose accuracy <65% correct (~15% exclusion) yielded consistent results throughout all our assessments. Assessing the consistency of performance throughout the task by measuring the

variability of the mean accuracy in windows of 60 trials, resulted in exclusion of only 8 more participants, whose variance exceeded the threshold of 4 standard deviations that we set. The number of participants before and after exclusions in each experiment is summarized in **Supplementary Table S2**.

This rate of exclusion is largely an outcome of our choice of a non-adaptive protocol. Few participants had poor performance even though they understood the task and could perform for very large frequency differences in the demo stage. We chose a non-adaptive protocol since adaptive protocols introduce correlations between the parameters of consecutive trials, which we aimed to avoid. In a non-adaptive protocol, the difference between the tones of a given trial ($\delta^t$) is always sampled from the same distribution, regardless of the participant's performance, ensuring that similar stimuli statistics are presented across participants.
The sample size for Experiment 1 was large, with 30 participants in each group, performing 600 trials each. Group size was determined by earlier simulations that assessed the required number of participants and trials (~25 participants). Experiment 2 aimed to recruit groups of individuals of a similar size. The M-Turk data were used to determine the required number of participants for Experiment 3. We administered 4 distributions (uniform of 3 octaves, uniform of 2 octaves, Gaussian and bimodal) and determined which is most informative with the smallest number of participants. We randomly sampled groups of 15 participants from the pool of participants in each distribution (~120) and found that in most 15-participant samples, we could calculate reliable bias functions for the bimodal distribution.

Assessments

All our participants performed 2-tone frequency discrimination. Lab participants (Experiments 1-3) were also administered a set of cognitive assessments, which evaluated

general reasoning skills by the standard Block Design task (WAIS-IV[42]), and reading abilities by pseudo-word and paragraph reading (described in ref 18). Results of these assessments are reported in **Supplementary Tables S1** and **S3** for Experiments 1-2, and Experiment 3, respectively.

The 2-tone discrimination task and protocols

In all 2-tone discrimination experiments, the design of the 2-tone discrimination task's basic trial structure and protocol were similar. The specific parameters differed mildly between experiments, but were always exactly the same for test and neurotypical groups. On each trial, participants heard two serially presented tones and were asked to determine which tone had the higher pitch and respond accordingly. In all experiments, a short demo session of 10 trials preceded the actual experiment. Feedback was provided during the demo; 80% success on the 10 demo (easy) trials was a prerequisite for continuing to the main task. In cases of failure to reach this success threshold, the training was repeated until the criterion was met.

*1. Experiment 1 – dyslexia and neurotypical groups*

Inter-stimulus interval (ISI) was 600ms, tone duration was 50ms and the delay from the participant's response to the onset of the first tone of the following trial (inter-trial interval; ITI) was 650ms. The frequency separation between the two tones was sampled log-uniformly from the range of 1–30% (i.e. uniformly from the range of 17.2–454.2 cents). The task was administered with a set of constant stimuli designed specifically for this experiment (available at: https://goo.gl/UnrG1A). This sequence was roughly Gaussian[18] with frequencies ranging between 550 and 1,800Hz (1.7 octaves). Each participant performed 600 trials. No feedback was provided for trials after successful completion of the demo. Both our pilot in-lab experiments and our M-Turk experiment reported here were conducted with

feedback (one of the M-Turk experiment's distributions was Gaussian). The results of these experiments did not differ from neurotypicals' performance in Experiment 1, indicating that the presence of feedback does not affect the pattern of participants' bias.

*2. Experiment 2 – ASD and neurotypical groups*

The ISI was 500ms, tone duration was 50ms and ITI was 1000ms. The frequency separation between the two tones was sampled log-uniformly from the range of 0.3–10% (uniformly from 5.19–165.0 cents). Tones were sampled log-uniformly between 800 and 1,250Hz (65% of an octave). Each participant performed 200 trials. Feedback was provided after every trial.

*3. Experiment 3 – ASD, dyslexia and neurotypical groups*

The ISI was 850ms, tone duration was 120ms and ITI was 650ms. In 90% of the block trials, the proportional frequency separations between the two tones were sampled log-uniformly from the range of 0.5-20% (uniformly from 8.63–315.6 cents). In the remaining 10%, the two tones were identical, making the task "impossible". These trials were included to increase the proportion of bias related information that we could extract from the participants' responses. Tones were sampled from a mixture of two non-overlapping log-uniform distributions with frequency ranges of 440-800Hz and 1,228-2,295Hz (0.9 of an octave spanned by each range). On a given trial, one of the modes was selected with equal probability, and the first tone was then sampled log-uniformly from that mode. Each participant performed 300 trials. Feedback was provided after every trial, and randomized in the "impossible" cases.

Both experimenters and participants were blind to the purpose of the study. Importantly, groups were not expected to and indeed generally did not, differ in main effects (i.e. similar perceptual thresholds). Hence, expectations and enthusiasm were not relevant to the parameters we compared.

*4. Amazon Mechanical Turk experiments*

Four 2-tone discrimination experiments were conducted on-line via the Amazon Mechanical Turk (M-Turk) crowed sourcing platform. The experiments were constructed in JavaScript and administered using web-browsers with tones that were preloaded and played using HTML 5 Audio. The temporal structure of each trial was the same as in Experiments 1-3 (illustrated in **Fig. 1a**). In each of the four experiments (Broad uniform, Narrow uniform, Gaussian and a Bimodal of two uniform distributions; see **Supplementary Table S2**), all basic stimuli parameters were identical: the inter-stimulus interval (ISI) within a trial was 840ms, tone duration was 120ms, and ITI was 650ms. The absolute difference between the two tones' log-frequencies was sampled uniformly from the range of 0.007-0.14 octaves (0.5% - 10% difference in Hz). The 2$^{nd}$ tone was either higher or lower than the 1$^{st}$ tone, with equal probability. The four experiments differed only in how the first tone of each trial was sampled, as described in **Supplementary Table S2**.

The assessment protocol was consistent of the following:

Familiarization with **pitch** → Familiarization with **task** → Task: **block 1** → **block 2** → **block 3.**

For familiarization with the concept of pitch, participants could click on any of 4 buttons, which resulted in playing 660, 780, 1,150 and 1,520Hz tones, respectively. Participants were asked to set the loudness to a comfortable level. Next, the participants performed a short training session to get familiar with the structure and interface of the task. They were given the following instructions: *"On each trial (question), two tones will be played consecutively: tone 1 → tone 2. Then, I ask: which of the two tones had the higher pitch*?"

Participants responded by either clicking on one of the two GUI buttons with their mouse, or by pressing a corresponding keyboard button. They were given visual (happy or sad smiley) feedback after every trial. Familiarization with the task was completed after 30 trials, or after 7 consecutive correct answers. It was followed by an optional short break, which could be initiated by the participant by clicking a GUI button. The main task consisted of 300 trials, performed in 3 blocks of 100 trials each. A short optional break followed each block. In addition to feedback after each trial, at the end of each block, the participants received information about their mean accuracy.

## Statistical analyses

Model fitting for both GLM and GAM was performed using the mgcv R package[43,44]. Hypothesis testing was performed using the scikit-learn and SciPy packages in Python. All tests used for hypothesis testing were two-sided. All comparisons were made using non-parametric test except in the case of the probit bias comparison, where we verified the assumption of normality using a Shapiro-Wilk normality test.

### Probit models

The probability that a participant reported hearing the frequency of the 2nd tone on trial $t$, $f_2^t$, as higher than that of the 1st , $f_1^t$, was modelled as:

$$p("f_2^t > f_1^t") = \Phi(\alpha_s^t \delta^t + b^t)$$

where $\Phi$ is the standard normal cumulative distribution function (i.e., $\Phi^{-1}$ is the probit function), $\alpha_s^t$ is a participant specific frequency sensitivity, which could change with time

during the assessment, $\delta^t$ is the difference of log-frequencies $f_2^t - f_1^t$ on the $t^{\text{th}}$ trial (all frequencies were expressed in logarithmic units throughout), and $b^t$ is a bias term for that trial that depends on stimulus history.

We explored three models, in increasing order of sophistication:

1) An aggregate model in which $\alpha_s^t$ does not depend on time, and $\delta^t = \pm\bar{\delta}$ and $b^t = \pm\bar{b}$ for constants $\bar{\delta}$ and $\bar{b}$.

2) A generalized linear model (GLM) in which $\delta^t$ is the actual log-frequency difference on the trial and $b^t = \sum_{i=1}^{4} w_i d_i^t + w_\infty d_\infty^t$ for linear weights $w_1, w_2, w_3, w_4, w_\infty$ associated with trial-specific history terms $d_1^t, d_2^t, d_3^t, d_4^t, d_\infty^t$ as defined in the main text.

3) A generalized additive model (GAM) in which $\delta^t$ was the frequency difference, and $b^t = b_1(d_1^t) + b_\infty(d_\infty^t)$ for non-parametric (spline-estimated) functions $b_1$ and $b_\infty$.

Average probit bias terms

We obtained a summary estimate of the magnitude of the contraction bias in each group. In our regression framework, each decision is modeled as driven by the objective task difficulty $\delta_t$ and a bias $b_t$ through the decision probability $p(y_t = 1|\delta_t + b_t) = \Phi(\alpha\delta_t + b_t)$.

Selecting a set of trials $T$, we can compute the average response $\langle y_t \rangle_T$. For this set of trials we can compute the average difficulty $\langle \delta_t \rangle_T$ , define an average bias $\langle b_t \rangle_T$ by $\langle y_t \rangle_T = \Phi(\alpha\langle \delta_t \rangle_T + \langle b_t \rangle_T)$, or the equivalent $\langle b_t \rangle_T = \Phi^{-1}(\langle y_t \rangle_T) - \alpha\langle \delta_t \rangle_T$.

Trials are grouped depending on whether $f_1 > f_2$ and whether $f_1 > \langle f \rangle$, yielding 4 trial types, where mean responses are expressed as

$$\langle y_t \rangle_T \begin{cases} \Phi(\alpha\delta + b) & : \mathrm{T}_1 = \{f_1 > f_2, f_1 > \langle f \rangle\} \\ \Phi(\alpha\delta - b) & : \mathrm{T}_2 = \{f_1 > f_2, f_1 < \langle f \rangle\} \\ \Phi(-\alpha\delta + b) & : \mathrm{T}_3 = \{f_1 < f_2, f_1 > \langle f \rangle\} \\ \Phi(-\alpha\delta - b) & : \mathrm{T}_4 = \{f_1 < f_2, f_1 < \langle f \rangle\} \end{cases}$$

where $b$ and $\delta$ denote a shared averaged bias and difficulty, respectively.

An estimator of the bias $b$ that does not require estimating participant's precision $\alpha$ can be derived from average response in each trial region:

$$\hat{b} = \frac{\Phi^{-1}\left(\langle y_t \rangle_{T_1}\right) + \Phi^{-1}\left(\langle y_t \rangle_{T_3}\right) - \Phi^{-1}\left(\langle y_t \rangle_{T_2}\right) - \Phi^{-1}\left(\langle y_t \rangle_{T_4}\right)}{4}$$

GLM Regression

The Generalized Linear Model has the following form (calculating separate parameters for 4 trials back):

$$p("f_2^{\mathrm{t}} > f_1^{\mathrm{t}}") = \Phi(\alpha_s^t \delta^t + \textstyle\sum_{i=1}^{4} w_i d_i^t + w_\infty d_\infty^t),$$

We fitted a single model to all the participants in each group, in which the weights $w_1, w_2, w_3, w_4, w_\infty$ determined the shared history-dependent bias, but the time-varying sensitivities $\alpha_s^t$ were participant-specific. In all analyses, responses in the first 5 trials were excluded.

GAM Regression

The Generalized Additive Model had the following form:

$$p("f_2^{\mathrm{t}} > f_1^{\mathrm{t}}") = \Phi\left(\alpha_s^t \delta^t + b_0 + b_1(d_1^t) + b_\infty(d_\infty^t)\right).$$

The constant offset $b_0$ was included to ensure that the learned bias functions $b_1$ and $b_\infty$ had zero means. However, it was never found to be significantly different from 0 and so is omitted when the model is discussed elsewhere in the text.

Varying-coefficient models and lapse rate

As with the GLM, a single model was fit to each group of participants, with common bias functions, but with participant specific sensitivities $\alpha_s$. We combined the pre-fitted $\alpha$-participant pairing in the model using the varying-coefficient models[45] method. This method assumes linearity in the regressors, but their coefficients are allowed to change smoothly with the value of other variable (alphas in our case): $g(\alpha^s)\delta_s^t$, Here $\alpha^s$ is the pre-fitted sensitivity parameter, $\delta_s^t$ is the difference between the target tones for each participant $s$, and $g$ is some smooth fitted function. In both GLM and GAM regression models we added a lapse parameter $\lambda$ to account for occasional inattentivness[46]: $p("f_2^t > f_1^t") = \frac{1}{2}\lambda + (1-\lambda)\phi$. We set $\lambda$ to a fixed 0.05 for all human performancers. IO model (described below) had $\lambda = 0$.

Random effects

In our regression analyses, GAM and GLM described above, we assumed shared bias functions across participants. This assumption is a statistical necessity, since we do not have sufficent individual data to estimate the model individually. To fit single participants, we assumed a shared parameter across participants, but allowed small individual deviations (termed random effects). Because assumed deviations are small, inference in these models is tractable.

Cross-validations

The quality of model fit was assessed by 10-fold cross-validation of the predictive accuracy. Trials were partitioned randomly into 10 validation subsets, sampled evenly across participants. Prediction was assessed using AUC: the area under the receiver operating characteristic (ROC) curve defined by comparing the model output to a swept threshold. Comparisons were tested for significance using a Wilcoxon Signed-rank test over cross-

validation fold scores. All cross-validations were performed with the "scikit-learn" package in Python, using the same random state.

Re-sampling methods

Testing for the significance of inter-group differences was based on a permutation test. Both GLM and GAM models were fit to 50,000 surrogate data sets obtained by permuting individual participant labels (preserving all data associated with that individual). This permutation procedure generated a null distribution against which the true difference in parameters or contributions to variance of bias between the groups could be compared to obtain a p-value.

Ideal Observer model

A normative account of the contraction bias has been proposed by Ashourian et al[16]. This account relies on a particular model of participants: both sensation and memory retention (for the first tone) are noisy and participants are aware of the statistics of these noises. On any given trial, tone frequencies $f_1$ and $f_2$ need to be inferred from their noisy representations as they are not directly observed and the decision on whether $f_2 > f_1$ is uncertain. One way to reduce uncertainty and increase decision accuracy is to incorporate statistical knowledge about the stimulus distribution. This can be formalized using the framework of probabilistic inference as follows.

Let $p(f)$ be the stimulus distribution assumed to be known to the participant. It is assumed that participants encode a noisy representation of the presented tones $\tilde{f_1}$ of $f_1$ and $\tilde{f_2}$ of $f_2$. In addition, by the time of decision the first tone representation is corrupted by additional memory noise (all noises are assumed to be independent).

These noise models provide likelihoods for the underlying value of tone frequencies $p(\tilde{f}_i|f_i)$. Under this model, an IO uses the stimulus distribution $p(f)$ as a prior distribution to reduce the uncertainty in the frequency of each tone by computing the posterior distribution over tones $p(f_i|\tilde{f}_i) \propto p(f_i)\, p(\tilde{f}_i|f_i)$.

Thus, we model each noise distribution as Gaussian in log-frequency, centered on the true value, $p(\tilde{f}_i|f_i) \sim \mathcal{N}(f_i, \sigma_i^2)$, with $\sigma_2^2 = \sigma_s^2$ reflecting sensory noise alone, but $\sigma_1^2 = \sigma_2^2 + \sigma_m^2$ incorporating additional noise associated with working memory.

Based on two noisy encoded tone frequencies $\tilde{f}_1, \tilde{f}_2$ the optimal decision on whether "$f_2 > f_1$" is given by:

$$\mathbf{1}[p(\text{"}f_2 > f_1\text{"}|\tilde{f}_1, \tilde{f}_2) > \tfrac{1}{2}],$$

where $\mathbf{1}$ is the truth function which equals 1 if the argument is true and 0 otherwise. This optimal decision can be rewritten as $\mathbf{1}[\hat{m}(\tilde{f}_1, \tilde{f}_2) > 0]$ where $\hat{m}(\tilde{f}_1, \tilde{f}_2) = \boldsymbol{median}[p(f_2 - f_1|\tilde{f}_1, \tilde{f}_2)]$ (see Sheppard[47]). In other words, an ideal decision will test whether the median of the posterior distribution on the frequency difference $f_2 - f_1$ exceeds 0. This decision rule gives a deterministic result based on the internally encoded values $\tilde{f}_i$.

For a trial on which a pair of tones with true frequencies $f_1, f_2$, is presented, the IO decision will vary with noise. We can compute the expected IO decision probability by averaging over different realizations of the encoding noise:

$$p(\text{"}f_2 > f_1\text{"}|f_1, f_2) = \boldsymbol{E}_{\tilde{f}_1, \tilde{f}_2 \sim p(\tilde{f}_1, \tilde{f}_2|f_1, f_2)} \mathbf{1}[\hat{m}(\tilde{f}_1, \tilde{f}_2) > 0]$$

This probability depends on the distribution of the median $\hat{m}(\tilde{f}_1, \tilde{f}_2)$ as $\tilde{f}_1$ and $\tilde{f}_2$ vary. We approximated this distribution by a Gaussian with matched mean and variance $\mathcal{N}(\hat{\mu}(f_1, f_2), \hat{\sigma}^2(f_1, f_2))$ where $\hat{\mu}(f_1, f_2) = E[\hat{m}|f_1, f_2]$ and $\hat{\sigma}^2(f_1, f_2) = Var[\hat{m}|f_1, f_2]$.

Using this approximation leads to an analytical expression for the decision probability which resembles the non-linear probit structure of the GAM model:

$$p("f_2 > f_1"|f_1, f_2) \approx \Phi\left(\frac{\hat{\mu}(f_1,f_2)}{\hat{\sigma}(f_1,f_2)}\right) = \Phi(\frac{f_2-f_1}{\hat{\sigma}(f_1,f_2)} + \hat{b}(f_1, f_2))$$

With $\hat{b}(f_1, f_2) = \frac{\hat{\mu}(f_1,f_2)-(f_2-f_1)}{\hat{\sigma}(f_1,f_2)}$ in particular, for "impossible" trials on which $f_1 = f_2$, we have $\hat{b}(f_1, f_2 = f_1) = \frac{\hat{\mu}(f_1,f_2)}{\hat{\sigma}(f_1,f_2)}$.

A Life Sciences Reporting Summary for this paper is available.

**Data availability**

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Additional references:**


42. Weschler, D. Wechsler Adult Intelligence Scale - Fourth Edition . 1–3 (2008).

43. Wood, A. S. & Wood, M. S. Package 'mgcv' Title Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. (2018).

44. Knoblauch, K. psyphy: Functions for analyzing psychophysical data in R. (2007).

45. Trevor, H. & Tibshirani, R. Varying-coefficients Models. *J. R. Stat. Soc.* **55,** 757–796 (1993).

46. Micheyl, C. Psychometric functions for pure-tone frequency discrimination. *Acoust. Soc. Am.* **130,** 263–272 (2011).

47. Sheppard, W. F. On the application of the theory of error to cases of normal distribution and normal correlation. *Philos. Trans. R. Soc. London* **192,** 101–531 (1899).