

1

A Computational Model of Dyslexics' Perceptual Difficulties as Impaired Inference of Sound Statistics

Sagi Jaffe-Dax, Ofri Raviv, Yonatan Loewenstein, and Merav Ahissar

The Hebrew University of Jerusalem, Jerusalem, Israel

Introduction—Contraction Bias in Simple Discrimination Tasks

Perception is a complex cognitive process in which noisy signals are extracted from the environment and interpreted. It is generally believed that perceptual resolution is limited by internal noise that constrains people's ability to differentiate physically similar stimuli. The magnitude of this internal noise is typically estimated using the two-alternative forced choice (2AFC) paradigm, which was introduced to eliminate participants' perceptual and response biases during experiments (Green & Swets, 1966; Macmillan & Creelman, 2004). In this paradigm, a participant is presented with two temporally separated stimuli that differ along a physical dimension and is instructed to compare them. The common assumption is that the probability of a correct response is determined by the physical difference between the two stimuli, relative to the level of internal noise. Performance is typically characterized by the threshold of discrimination, referred to as the Just Noticeable Difference (JND). Thus, the JND is a measure of the level of internal noise such that the higher the JND, the higher the inferred internal noise.

However, if the stimuli are highly predictable, perceptual resolution may not be limited by the magnitude of the internal noise. In other words, the assumption of a one-to-one

correspondence between the JND and the internal noise may ignore this potential benefit that derives from previous experience. If the internal representation of a stimulus is noisy and hence unreliable, prior expectations should bias the participant against unlikely stimuli. The larger the uncertainty of the measurements, the larger the contribution of these prior expectations is likely to be. The Bayesian theory of inference defines computationally how expectations regarding the probability distribution of stimuli should be combined with the noisy representations of these stimuli in order to form an optimal posterior percept (Knill & Richards, 1996).

One limitation of the Bayesian model is that it relies heavily on the assumption that the prior distribution of stimuli is known to the observer. While this assumption may be plausible in very long experiments comprising a large number of trials (e.g., thousands in Körding & Wolpert, 2004) or in experiments utilizing natural tasks (e.g., in reading; Norris, 2006), it is unclear to what extent a rich Bayesian inference is formed when participants have less experience with a task.

Here, we studied participants' patterns of responses on a 2AFC tone discrimination task in relatively short experiments consisting of tens of trials. We found a substantial context effect, whose extent depended on the statistics of the stimuli used in the task and on participants' internal noise level. Participants' pattern of behavior was consistent with an "implicit memory" model in

which the representation of previous stimuli is a single scalar that continuously updates with examples. Thus, this model can be viewed as a simple implementation of the Bayesian model that provides a better account of participants' perceptual decision making. We then applied this model to a special population of dyslexic subjects and found that this model captures their difficulties on such tasks.

Contraction Bias—a Simple Experimental Measure of Context Effects

In order to evaluate the impact of the stimulus statistics on perception parametrically, we used a 2AFC frequency discrimination task. On each trial, participants were sequentially presented with two pure tones and instructed to indicate which had a higher pitch (illustrated in Fig. 1.1). The mean frequency of each pair was uniformly selected from a broad range and the frequency difference was chosen either adaptively or according to a pretesting decision. We termed this protocol the *No-Reference*, since it differs from typical psychophysical assessments where one of the two stimuli on each trial serves as a reference and repeats across trials. Though frequency discrimination tasks are traditionally used as an assessment of low-level sensory bottlenecks, we have shown that performance is highly affected by context, both in the No-reference protocol (Raviv, Ahissar, & Loewenstein, 2012) and in the various reference protocols. In fact the form of integration of previous stimuli explains seemingly inconsistent biases in success rate depending on the position of the reference stimulus within the trial (Raviv, Lieder, Loewenstein, & Ahissar, 2014).

Specifically, expectations, formalized as the prior distribution of the stimuli used in the experiment, have been shown to bias participants' responses in a way that is often (though not always) consistent with the Bayesian framework (reviewed in Körding, 2007). In particular, responses in the 2AFC paradigm

have been shown to be biased by prior expectations. Thus, when the magnitude of the two stimuli is small with respect to the mean of the previous stimuli used in the experiment, participants tend to respond that the second stimulus was smaller, whereas when the magnitude of both stimuli is large they tend to respond that the second stimulus was larger (Preuschhof, Schubert, Villringer, & Heekeren, 2010; Woodrow, 1933). We have shown that this bias, known as the “contraction bias,” can be understood within the Bayesian framework. Rather than comparing the two noisy representations of the stimuli, the participant combines the noisy representations of the two stimuli with the prior distribution of the stimuli to form two posterior distributions. The two posteriors are compared to maximize the probability of a correct response. The contribution of the prior distribution to the two posteriors is not equal. The larger the level of noise in the representation of the stimulus, the larger the contribution of the prior distribution to the posterior (Ashourian & Loewenstein, 2011). The level of noise in the representation of the first stimulus is larger than the level of noise in the representation of the second stimulus because of the additional noise associated with the encoding, and maintenance of the first stimulus in memory during the inter-stimulus interval of sequential presentation tasks (Bull & Cuddy, 1972; Wickergren, 1969). As a result, the posterior distribution of the first stimulus is biased more by the prior distribution than the posterior distribution of the second stimulus. Since the posterior of the first stimulus is contracted more than the posterior of the second stimulus, participants' responses are biased toward overestimating the first stimulus when it is small and underestimating it when it is large with respect to the prior (distribution of previous stimuli).

The expected outcome of the contraction bias on performance is that the combination of the mean frequency on the trial with respect to the mean frequency of the experiment, and the relative frequency of the two tones on the trial determines the impact of experiment's statistics in the following manner: *Bias+* trials

are trials in which the experiment's statistics is expected to improve performance. Specifically, a stronger "pulling" of the first (compared to the second) tone toward the average frequency increases the difference between the representations of the two tones in the trial. In such trials (1) the frequencies of the two tones are either both higher than the mean frequency (in log scale) or both lower than the mean frequency, and (2) the frequency of the second tone is more extreme than that of the first tone (yellow zones in Fig. 1.1). By contrast, *Bias*-trials are trials in which the statistics are expected to hamper performance. Specifically, when the two tones are above the mean, and the second tone is lower than the first, and when they are lower than the mean, and the second tone is higher than the first (gray zones in Fig. 1.1). In these trials, contracting the first tone toward the mean frequency decreases its perceived difference with the second tone and is thus expected to be detrimental to performance. *Bias0* trials (white zones in Fig. 1.1) are trials in which the first and the second tone flank the mean frequency.

Dyslexia

Dyslexia is defined as a "specific and significant impairment in the development of reading skills that is not accounted for by mental age, visual acuity problems, or inadequate schooling" (WHO, 2010) and affects 5% of the world's population (Lindgren, De Renzi, & Richman, 1985). The standard assumption is that dyslexics' phonological representations, which need to be accurate for efficient usage of the alphabetical code, are impaired (Snowling, 2000). However, dyslexics perform well on some tasks that rely on phonological representations (reviewed in Ramus & Ahissar, 2012). Moreover, dyslexics perform poorly on many simple auditory tasks that have no phonological component. In experimental settings, more than half of the tested dyslexic participants performed poorly on two-tone frequency discrimination batteries, and this performance was correlated with verbal memory skills (Banai & Ahissar, 2004; McArthur and Bishop, 2004).

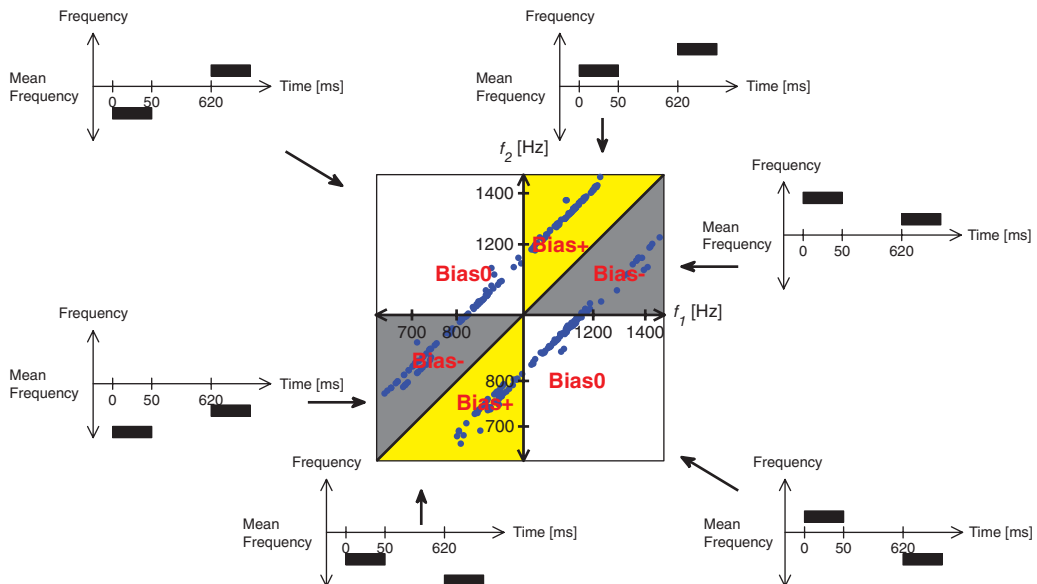


Figure 1.1 Illustration of trials' distribution and the contraction bias. Middle: Trials' distribution on the $[f_1, f_2]$ plane. Each dot denotes f_1 and f_2 of a trial. Surrounding: schematic examples of the three types of trials. In *Bias+* trials the first tone is closer to the mean frequency. In *Bias-* trials the first tone is farther from the mean. In *Bias0* trials the two tones flank the mean.

Banai and Ahissar (2006) aimed to pinpoint the bottleneck to this performance by testing task protocols that relied to a greater or a lesser extent on implicit memory of previous trials and only found a deficit in the former. Ahissar et al. (2006) tested this reduced sensitivity to recently presented stimuli and found a similar phenomenon for speech stimuli. Based on these observations they proposed the anchoring deficit hypothesis (Ahissar, 2007; Ahissar, Lubin, Putter-Katz, & Banai, 2007), which posits that dyslexics' deficits stem from poor utilization of stimulus repetitions. Thus, dyslexia does not result from a deficit in stimulus processing (sensation) but from a lesser ability to process information from previous presentations (similar to the concept of predictive coding; Díaz, Hintz, Kiebel, & von Kriegstein, 2012).

The Magnitude of Contraction Bias is Smaller in Dyslexics than in Controls

Raviv et al. (2012) measured the magnitude of the contraction bias (the difference in success rate between *Bias+* and *Bias-* trials, as illustrated in Fig. 1.1) in the general population in the No-Ref protocol and found a substantial effect. Even larger context effects in the general population of good readers (students) were replicated by Jaffe-Dax et al. (2015). They used a relatively fixed frequency difference that was previously found to yield an average of 80% correct across trials. Though the difficulty of each trial was fixed (in terms of inter-stimuli frequency difference), average performance differed tremendously with respect to the tone distributions (Fig. 1.2A). Interestingly, in the *Bias-* range, the average performance of the controls did not significantly differ from chance (Wilcoxon test, $p = .55$).

Based on the Anchoring Deficit hypothesis we inquired whether dyslexics' difficulties in benefitting from simple stimulus

repetitions also applied to the more general summary statistics of the experiments. Specifically, we examined whether dyslexics' context effects were reduced even when there was no reference, and the mean frequency was retrieved by some form of integration across previous trials. To test this, a group of dyslexics matched for age, education and general reasoning skills was also tested on the same 2AFC frequency discrimination protocol. Overall, dyslexics performed more poorly than the controls. However, they showed a smaller context effect; that is, a smaller difference in performance between *Bias+* and *Bias-* trials (Fig. 1.2B). Hence importantly, in spite of their overall lower performance, dyslexics' performance in the *Bias-* region was significantly above chance (58.2% Wilcoxon test, $p < .05$), whereas controls' performance was at the chance level.

The Implicit Memory Model (IMM) Account for the Contraction Bias

Raviv et al. (2012) derived a simple model of implicit memory to account for the contraction effect. We used to the same model to analyze the behavior of controls and dyslexics, and test for putative differences between them. According to this model, choices on each trial are determined by the difference between the frequency of the second tone and a weighted average (M_1) of the noisy memory of present and past frequencies of the first tone. Formally,

$$a(t) = \text{sgn}(M_1(t) - f_2(t)) \quad (\text{Eq. 1.1})$$

where $a(t)$ denotes the choice of the participant on trial t , sgn is the sign function, $f_2(t)$ is the frequency of the second tone on trial t , and:

$$M_1(t) = \eta M_1(t-1) + (1-\eta)(f_1(t) + z(t)) \quad (\text{Eq. 1.2})$$

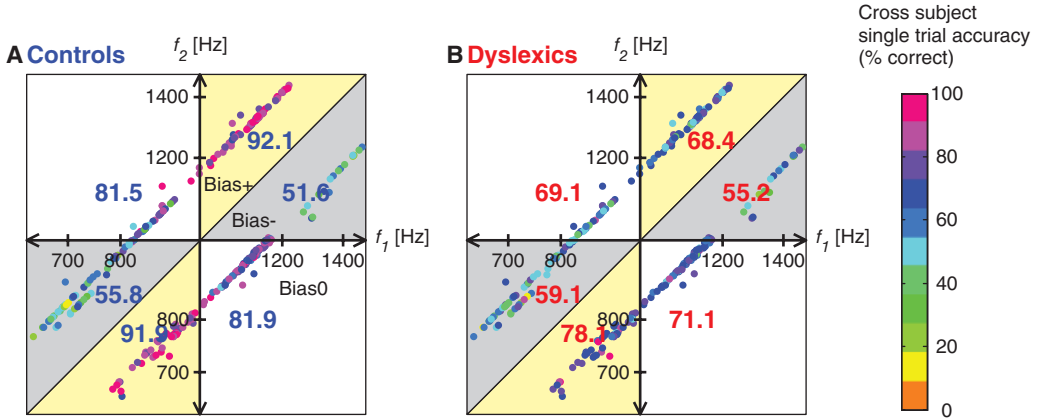


Figure 1.2 Performance of controls and dyslexics differentially depend on trial type. Mean performance (% correct) of controls (A) and dyslexics (B) in the six subregions of trial types, plotted on the frequency plane of the second tone f_2 as a function of the first tone f_1 . *Bias+* zones (denoted in yellow) are above the diagonal when both tones are above the mean frequency (second tone is higher) and below the diagonal when both tones are below the mean frequency (second tone is lower). *Bias-* regions (denoted in gray) are complementary with respect to the diagonal, and *Bias0* trials (denoted in white) are those trials associated with the two remaining quarters. Each dot denotes f_1 and f_2 of a trial (tested across individuals). The color of each dot denotes the cross-subject average performance for that pair of stimuli. Numbers denote the average percent correct in each subregion.

where $f_1(t)$ is the frequency of the first tone on the trial, $z(t)$ is an independent Gaussian random variable with zero mean and variance σ^2 , and we assume resetting of the initial conditions $M_1(t=1) = f_1(t=1) + z(t=1)$ (Shteingart, Neiman, & Loewenstein, 2013). This model is characterized by two parameters, σ and η . The first parameter, σ , denotes the level of internal noise in the process of “sensing and memorizing” the first tone on each trial. The second parameter, η , denotes the weight of previously stored stimuli in the current comparison.

Therefore, the probability of responding “first tone higher” is:

$$\Pr["1"] = \Phi\left(\frac{M_1(t) - f_2(t)}{\sigma_e(t)}\right) \quad (\text{Eq. 1.3})$$

where Φ is the cumulative normal distribution function and $\sigma_e(t)$ is the effective response variance at trial t , which depends on σ , η , and t :

$$\sigma_e = \sigma \sqrt{\eta^{2t-2} + (1 - \eta^{2t-2}) \frac{1 - \eta}{1 + \eta}} \quad (\text{Eq. 1.4})$$

For each participant, we estimated the two parameters, η and σ , that minimize the squared distance between the predicted response probabilities of the model and the observed responses in the frequency discrimination task. To assess the reliability of this estimate, we bootstrapped the trials for each participant by 1,000 samples of 300 trials with replacement and re-estimating parameters.

We then estimated the optimal impact of implicit memory given this estimated σ ; that is, for each participant we found η^* – the η that would maximize success on the auditory task. We defined inadequacy, or suboptimality, of the implicit memory weighting of each participant as the difference between η^* and the estimated η .

The model (Eq. 1.2) is characterized by two parameters: η , defined as the contribution of previous trials ($\eta = 0$, no contribution; the larger the value of η , the larger the contribution), and σ , the level of internal noise (on a scale of percentage difference between the two tones: the larger the value of σ the

noisier the within-trial representation). We used the IMM to estimate the values of η and σ for each participant. We tested the model by simulating it on the same task using the estimated parameters. Qualitatively, the differential performance on the three trial types and the different performance level of dyslexics and controls was captured by the model (Fig. 1.3). Quantitatively, the difference in performance between the *Bias+* and *Bias-* regions was slightly larger in the experiment than in the model, in particular for the control participants.

One prediction of the Bayesian framework is that the effect of the stimulus statistics on behavior should increase when the level of internal noise increases. Qualitatively, this prediction is intuitive. If the representation of the stimuli is noiseless, performance cannot be improved by incorporating prior information. However, if the representation of a stimulus is noisy, prior information should be useful, and the noisier the representation, the larger the weight that should be given to this prior on the discrimination task. One study in fact showed that in the visual

modality, increasing the level of internal noise enhanced the contribution of prior knowledge to perception and decision making (Ashourian & Loewenstein, 2011).

The IMM makes no assumptions regarding the relationships between participants' η and σ values. Nevertheless, the model can be used to determine the extent to which participants' weighting of previous trials (their η value) was close to optimal given their within-trial noise (σ) and the stimulus statistics (in the sense of maximizing their success rate; Fig. 1.4A, green).

Analysis of the parameters characterizing our participants indicated that, on average, the estimated value of σ was higher among the dyslexic participants (median (IQR), controls: $\sigma = 35$ (40)%; dyslexics: $\sigma = 98$ (213)%; Mann-Whitney test, $z = 2.2$, $p < .05$). Given their higher σ value, the optimal model solution predicted that the dyslexics' η should also be higher than the controls' η value. However, the dyslexics' η value was similar to that of the controls (median (IQR), controls: $\eta = .52$ (.5); dyslexics: $\eta = .41$ (0.53); Mann-Whitney test, n.s.).

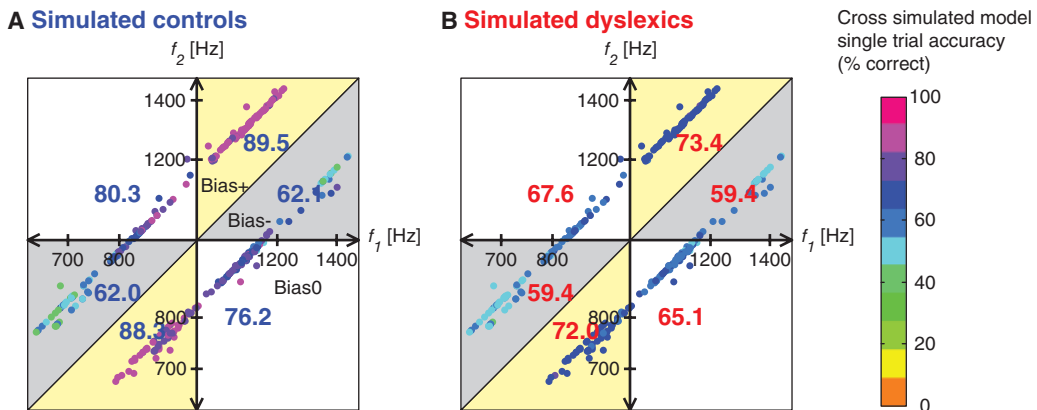


Figure 1.3 Simulating the experiment with the IMM produces similar results to those measured experimentally (shown in Fig. 1.2). (A, B) Mean performance (% correct) of simulated controls (A) and simulated dyslexics (B) in the six subregions of trial types plotted on the $[f_1, f_2]$ plane. *Bias+* zones are denoted in yellow, *Bias-* in gray, and *Bias0* in white. Each dot denotes f_1 and f_2 of a trial (tested across simulated individuals). The color of each dot denotes the average cross-simulation performance for that pair of stimuli. Numbers correspond to the percentage correct in each sub-region.

Dyslexics Underweight Previous Trials Given Their Internal Noise Level

The IMM model was used to assess whether dyslexics underweight previous trials given their (calculated) level of internal noise. We calculated the optimal η value, that is, the value of η that minimizes the probability of an error in the stimulation schedule for a fixed value of σ .

We defined mental difference ($D(t)$) and correct difference ($C(t)$) on each trial as:

$$D(t) = f_2(t) - M_1(t) \quad (\text{Eq. 1.5})$$

$$C(t) = f_2(t) - f_1(t) \quad (\text{Eq. 1.6})$$

Thus, the probability of making a correct decision on each trial is:

$$\begin{aligned} & \Pr[\text{correct}] \\ &= \Pr[\text{sign}(D(t)) = \text{sign}(C(t))] \quad (\text{Eq. 1.7}) \end{aligned}$$

According to Sheppard's Median Dichotomy Theorem (Kendall, Stuart, & Ord, 1987, p. 482; Sheppard, 1899):

$$\begin{aligned} & \Pr[\text{sign}(D(t)) = \text{sign}(C(t))] \\ &= \frac{1}{2} + \frac{1}{\pi} \sin^{-1}(\rho_m) \quad (\text{Eq. 1.8}) \end{aligned}$$

where: $\rho_m = \frac{\text{cov}(D, C)}{\sqrt{\text{var}(D)\text{var}(C)}}$ is the correlation between the mental difference and the correct difference.

The probability of a correct response is a monotonic function of ρ_m over the relevant scope [0,1]; thus, it is sufficient to maximize ρ_m over η :

$$\rho_m = \frac{(2 - \eta + (\eta - 2)\rho_f)\sigma_f^2}{\sqrt{\left(\sigma_f^2 + 2\sigma_f^2(\eta - 1)\rho_f + \eta^2\left(\frac{1 - \eta}{1 + \eta}(\sigma_f^2 + \sigma^2)\right) + (1 - \eta)^2(\sigma_f^2 + \sigma^2)\right)}(2(1 - \rho_f)\sigma_f^2)} \quad (\text{Eq. 1.9})$$

where $\rho_f = \frac{\text{cov}(f_1, f_2)}{\sqrt{\text{var}(f_1)\text{var}(f_2)}}$ is the correlation between the two tones and σ_f^2 is the variance of f_1 (the overall variance of the frequencies of the first tones), which is also equivalent to the variance of f_2 (the same marginal distributions for the frequencies of the first and the second tones).

The optimal η is a solution to the equation:

$$\frac{\partial \rho_m}{\partial \eta} = 0 \quad (\text{Eq. 1.10})$$

The solution to this equation was found using Mathematica 9.0 (Wolfram Research, Inc., Champaign, IL).

Since this computation assumes an infinite number of trials, the solution is only an asymptotic approximation of the optimal η . In order to verify its proximity to the optimal η value, with a finite number of trials, we ran numerical simulations. We estimated the η value that yielded the highest accuracy rate in the actual sequence of trials used in the experiment, for various values of σ and compared the two solutions.

As shown in Fig. 1.4A, controls' η values (blue squares) were nearly optimal (green line) given their σ values, indicating that their weighting of the history was similar to that of an "ideal listener" (in the framework of the IMM). By contrast, dyslexics' η values were lower than expected given their σ values (red squares). To quantify this group difference, we calculated the difference between the optimal and the actual η value for each participant. We found that dyslexics' underweighting of implicit memory was significantly larger than the controls' (Fig. 1.4B; Mann-Whitney test, $z = 2.6$, $p < .01$). Hence, given

their internal noise level, dyslexics, in contrast to controls, do not give sufficient weight to prior information.

Taken together, these findings show that in the framework of the IMM, controls' weighting of past events was nearly optimal. By contrast, dyslexics' weighting of these past events was too low, indicating a deficit in adequate incorporation of prior knowledge into perception.

General Discussion

In this chapter we showed that the contraction bias is a dominant determinant of participants' behavior in a 2AFC tone frequency discrimination task. Though clearly evident also in the pattern of dyslexics' behavior, its magnitude is significantly smaller.

The IMM, Detection Theory, and the Bayesian Framework

Some features of the contraction bias (and of the IMM) are consistent with the behavior of an ideal detector that utilizes the prior distribution to maximize performance. What information does the cognitive system

store about the prior distribution? The full Bayesian model represents an extreme approach, in which it is assumed that the participant has full information about the joint distribution of the two stimuli. The standard way in which signal detection theory is applied to psychophysics represents the other extreme, in which the participant does not have (or does not utilize) any prior information about the identity of the stimuli (but only about the probability of each response being correct; Green & Swets, 1966). The contraction bias in Fig. 1.2 demonstrates that participants have some information about the marginal probabilities.

The IMM can be viewed as a minimal modification of the standard approach of applying signal detection theory to perception in the direction of the full Bayesian model. Here, participants represent the prior distribution of the stimuli with a single scalar, which is an estimate of the mean of the marginal of the prior distribution. Despite its simplicity, this implicit model captures many facets of the behavioral results. Further studies are needed to determine whether, and to what extent other moments of the prior distributions are learned and utilized in the 2AFC discrimination

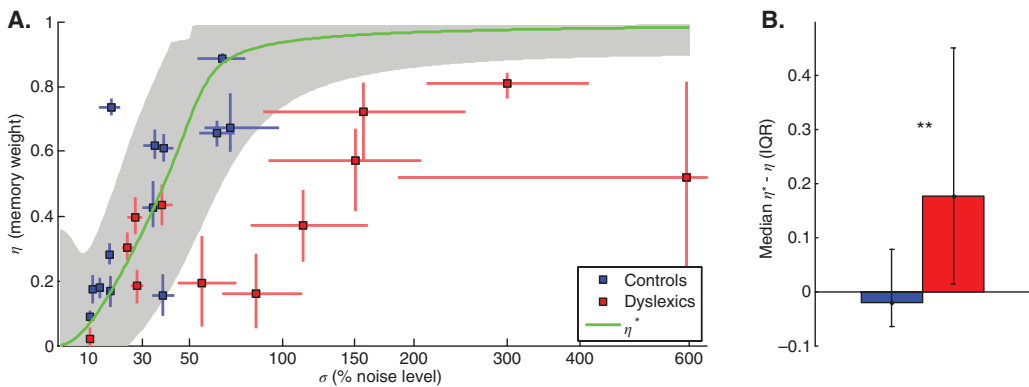


Figure 1.4 Estimated parameters of the Implicit Memory Model for controls and dyslexics. (A) Estimated values of η (which determine the proportion of implicit memory in the current representation of f_i) as a function of the estimated values of σ (percentage of internal noise with respect to the difference between the tones within the trial) of controls (blue; error bars denote bootstrapped inter-quartile range) and dyslexics (red). For all values of σ the optimal value of η (the value of η that minimizes the probability of an error in the experiment), denoted as η^* , is plotted in green. Gray area depicts the values of η that yield more than 97.5% of the optimal performance. (B) Median of the deviations from optimal weighting of previous trials. Dyslexics' deviation was significantly larger than controls' (Mann-Whitney test, $z = 2.5$, $p < .01$). Error bars denote inter-quartile range.

task, especially under longer exposure to distribution statistics.

Several studies have shown that the magnitude of the contribution of the prior distribution to perception on a given trial depends on the level of internal noise (Körding & Wolpert, 2004; Trommershauser et al., 2005). In particular, in the framework of the 2AFC task, increasing the delay between the first and second stimuli (Hanks et al., 2011; Lu, Williamson, & Kaufman, 1992) or introducing a distracting task between them (Ashourian & Loewenstein, 2011) enhances the contraction bias. These results are consistent with the Bayesian approach and can be accounted for in the framework of the IMM by optimal weighting of previous trials. In line with these studies, we found that controls' weighting of the mean of previous trials did not differ from optimal. However, dyslexics' weighting was significantly lower than optimal.

It should however be noted that our simple, two-parameter IMM only assumes within-trial noise. Although the dyslexics' underweighting of the stimulus statistics was suboptimal in this case, their weighting of history could be close to optimal if their deficit leads to increased memory noise. Introducing this to the model would require the addition of a third parameter, whereas our model accounts well for controls' performance with only two parameters. On the other hand, we could assume optimal weighting of the history (i.e., the weighting of the history is determined by the stimulus statistics) and replace η with a parameter that estimates the noise of previous trials (memory noise). In that case, increasing the weight of previous trials might not enhance their performance since it would also increase the memory noise. However, this would require an additional optimality assumption, which for this reason we did not pursue.

The Contribution of the Computational Model to Understanding Dyslexia

The IMM is an extended formalization of the anchoring deficit hypothesis of dyslexia (Ahissar, 2007; Ahissar et al., 2006),

which posits that repeated stimuli serve as an anchor, and thus boost performance when these stimuli are subsequently used. According to this theory, the ability to track such simple regularities is deficient in dyslexia. The IMM presented here proposes a specific, well-defined computation that is impaired in dyslexics' incorporation of stimulus statistics. The model specifies the conditions where incorporating previous trials is expected to improve perception and those where it is expected to hamper perception. It shows that stimulus repetition is not necessary (e.g., *Bias+* trials in a protocol with no reference), and may even not be beneficial (*Bias-* in a fixed reference protocol) for performance. The IMM predicts that dyslexics will only perform worse than controls on the trials that benefit from the stimuli statistics.

The anchoring deficit hypothesis and its computational implementation are inconsistent with the hypothesis that poor phonological representations are the core deficit in dyslexia (e.g., Snowling, 2000). By contrast, our computational description of dyslexics' difficulties is tightly related to hypotheses that associate dyslexics' difficulties with a failure to make effective predictions that can facilitate task performance ("predictive coding"; Díaz et al., 2012). It is also compatible with hypotheses that dyslexics are less resilient to external noise (the "noise exclusion hypothesis"; e.g., Conlon, Lilleskaret, Wright, & Power, 2012; Beattie, Lu, & Manis, 2011; Partanen et al., 2012; Sperling, Lu, Manis, & Seidenberg, 2005, 2006). According to the Bayesian framework underpinning the IMM, the prior information is utilized to compensate for the noise in the representation of the stimuli. We found that dyslexics do not properly adjust the weight of previous trials to the level of internal noise (Fig. 1.4). Functionally, this results in reduced noise exclusion. Importantly, putting this broad concept in a computational framework leads to a counterintuitive prediction: when the context is compromising (e.g. *Bias-* like conditions leading to disruptive predictions), dyslexics should not do worse and may even do better than controls since this mechanism for "noise exclusion" is biased by the prior statistics.

Other studies have suggested that dyslexics' implicit sequence learning is impaired (e.g., Jiménez-Fernández, Vaquero, Jiménez, & Defior, 2011; Stoodley, Harrison, & Stein, 2006; Vicari et al., 2003). This phrasing is very broad. For example, it is not clear whether the representation of syllables should be intact, whereas the representation of words, which are perhaps formed by implicit sequential learning (conditional probabilities between adjacent syllables; Saffran, Aslin, & Newport, 1996) should be impaired. Our computational model assumes poor incorporation of basic attributes (zero-order statistics), as observed by the poor usage of the mean frequency of the experiment. It predicts that the same deficit may lead to reduced sensitivity to the prevalence of single syllables.

Other studies (Boets et al., 2013; Ramus, 2014; Ramus & Szenkovits, 2008) have proposed that phonological representations remain intact, but that dyslexics' access to these representations is inefficient, perhaps due to impaired connectivity between the superior temporal areas that encode auditory stimuli and the frontal areas (e.g., Broca's region) which utilize them. This hypothesis assumes a clear distinction between representations and access, though it does not specify what "access" is, and whether it precedes perception. Consequently, it is too loosely defined, and does not specify which retrieval conditions are expected to pose difficulties

(e.g., whether implicit priming should be impaired) and whether they are expected to degrade perception or merely voluntary memory retrievals. Our own perspective, as described above, is that perception is never devoid of context (Raviv et al., 2014). Ease of retrieval is heavily affected by the availability of adequate predictions, which can substantially facilitate the process. Thus, difficulty in utilizing priors may impair the efficiency of retrieval. Nevertheless, the nature of retrieval processes in different behavioral contexts is far from being understood.

Recently, it has been shown that a hierarchical model of perception is able to reconstruct or decode the dynamics underlying a generated series of stimuli (Yildiz, von Kriegstein, & Kiebel, 2013). Our observations can be interpreted within this framework, and suggest that dyslexics' deficit resides in the higher and slower level of the hierarchy, which is characterized by larger time constants, or in their impaired communication with lower levels (Boets et al., 2013). The slower dynamics in this higher level are able to track cumulating evidence and form a prior based on the underlying statistics of the stimuli. In turn, these predictions are used to modulate the lower levels of the hierarchy to better match the upcoming stimuli and the requirements of the task. This coincides with data showing that among dyslexics, task-related top-down modulation is impaired (Díaz et al., 2012).

References

- Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Science*, 11(11), 458–465.
- Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, 9(12), 1558–1564.
- Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One*, 6(5), e19551.
- Banai, K., & Ahissar, M. (2004). Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiology and Neuro-Otology*, 9(6), 328–340.
- Banai, K., & Ahissar, M. (2006). Auditory processing deficits in dyslexia: Task or stimulus related? *Cerebral Cortex*, 16(12), 1718–1728.
- Beattie, R. L., Lu, Z.-L., & Manis, F. R. (2011). Dyslexic adults can learn from repeated

- stimulus presentation but have difficulties in excluding external noise. *PLoS One*, 6 (11), e27893.
- Boets, B., Op de Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., . . . & Ghesquière, P. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science*, 342(6163), 1251–1254.
- Bull, A. R., & Cuddy, L. L. (1972). Recognition memory for pitch of fixed and roving stimulus tones. *Perception and Psychophysics*, 11(1B), 105–109.
- Conlon, E. G., Lilleskaret, G., Wright, C. M., & Power, G. F. (2012). The influence of contrast on coherent motion processing in dyslexia. *Neuropsychologia*, 50(7), 1672–1681.
- Díaz, B., Hintz, F., Kiebel, S. J., & von Kriegstein, K. (2012). Dysfunction of the auditory thalamus in developmental dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34), 13841–13846.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley-Blackwell.
- Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E. & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, 31(17), 6339–6352.
- Jiménez-Fernández, G., Vaquero, J. M. M., Jiménez, L., & Defior, S. (2011). Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. *Annals of Dyslexia*, 61(1), 85–110.
- Kendall, M. G., Stuart, A., & Ord, J. K. (1987). *Advanced theory of statistics*. London: Charles Griffin.
- Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
- Körding, K. (2007). Decision theory: What “should” the nervous system do? *Science*, 318(October), 606–610.
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244–247.
- Lindgren, S. D., De Renzi, E., & Richman, L. C. (1985). Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Development*, 56(6), 1404–1417.
- Lu, Z.-L., Williamson, J., & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science*, 258(5088), 1668–1670.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user’s guide*. Abingdon: Taylor & Francis.
- McArthur, G. M., & Bishop, D. V. M. (2004). Which people with specific language impairment have auditory processing deficits? *Cognitive Neuropsychology*, 21(1), 79–94.
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychology Review*, 113(2), 327–357.
- Partanen, M., Fitzpatrick, K., Mädler, B., Edgell D., Bjornson B., & Giaschi D. E. (2012). Cortical basis for dichotic pitch perception in developmental dyslexia. *Brain and Language*, 123(2), 104–112.
- Preuschhof, C., Schubert, T., Villringer, A., & Heekeren, H. R. (2010). Prior information biases stimulus representations during vibrotactile decision making. *Journal of Cognitive Neuroscience*, 22(5), 875–887.
- Ramus, F. (2014). Neuroimaging sheds new light on the phonological deficit in dyslexia. *Trends in Cognitive Science*, (December 2013), 3–6.
- Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, 29(1–2), 37–41.
- Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology (Hove)*, 61(1), 129–141.
- Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How recent history affects perception: The normative approach and its heuristic approximation. *PLoS Computational Biology*, 8(10), e1002731.

- Raviv, O., Lieder, I., Loewenstein, Y., & Ahissar, M. (2014). Contradictory behavioral biases result from the influence of past stimuli on perception. *PLoS Computational Biology*, *10*(12), e1003948.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
- Sheppard, W. F. (1899) On the application of the theory of error to cases of normal distribution and normal correlation. *Philos. Trans. R. Soc. London*, *192*, 101–167.
- Shteingart, H., Neiman, T., & Loewenstein, Y. (2013). The role of first impression in operant learning. *Journal of Experimental Psychology: General*, *142*(2), 476–488.
- Snowling, M. J. (2000). *Dyslexia*. Oxford: Wiley-Blackwell.
- Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*, *8*(7), 862–863.
- Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2006). Motion-perception deficits and reading impairment: it's the noise, not the motion. *Psychological Science*, *17*(12), 1047–1053.
- Stoodley, C. J., Harrison, E. P. D., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, *44*(5), 795–798.
- Trommershauser, J., Gepshtein, S., Maloney, L. T., Landy, M. S., & Banks, M. S. (2005). Optimal compensation for changes in task-relevant movement variability. *Journal of Neuroscience*, *25*(31), 7169–7178.
- Vicari, S., Marotta, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia*, *41*(1), 108–114.
- WHO (World Health Organization) (2010). *International classification of diseases (ICD-10)*. Geneva: World Health Organization, F81.0.
- Wickergren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, *6*, 13–61.
- Woodrow, H. (1933). Weight-discrimination with a varying standard. *American Journal of Psychology*, *45*(3), 391–416.
- Yildiz, I. B., von Kriegstein, K., & Kiebel, S. J. (2013). From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLoS Computational Biology*, *9*(9), e1003219.