The mechanistic foundation of Weber's law

Jose L. Pardo-Vazquez, Juan R. Castiñeiras-de Saa1, Mafalda Valente, Iris Damião, Tiago Costa, M. Inês Vicente, André G. Mendonça, Zachary F. Mainen[1] and Alfonso Renart

Author information

Jose L. Pardo-Vazquez

Present address: Neuroscience and Motor Control Group, University of A Coruña, A Coruña, Spain

These authors contributed equally: Jose L. Pardo-Vazquez, Juan R. Castiñeiras-de Saa.

Affiliations

Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal

Jose L. Pardo-Vazquez, Juan R. Castiñeiras-de Saa, Mafalda Valente, Iris Damião, Tiago Costa, M. Inês Vicente, André G. Mendonça, Zachary F. Mainen & Alfonso Renart

Contributions

J.L.P.-V. and A.R. conceived the project. J.L.P.-V. and M.V. conducted the rat auditory experiments. I.D., M.V. and J.R.C.-S. conducted the human auditory experiments. M.I.V., A.G.M. and Z.F.M. conducted the rat olfactory experiments. A.R., J.R.C.-S. and T.C. conceived the theory. J.R.C.-S., J.L.P.-V., M.V. and A.R. analyzed the data. A.R. wrote the manuscript with feedback from all the authors.

Corresponding authors

Correspondence to Jose L. Pardo-Vazquez or Alfonso Renart.

Abstract

Although Weber's law is the most firmly established regularity in sensation, no principled way has been identified to choose between its many proposed explanations. We investigated Weber's law by training rats to discriminate the relative intensity of sounds at the two ears at various absolute levels. These experiments revealed the existence of a psychophysical regularity, which we term time–intensity equivalence in discrimination (TIED), describing how reaction times change as a function of absolute level. The TIED enables the mathematical specification of the computational basis of Weber's law, placing strict requirements on how stimulus intensity is encoded in the stochastic activity of sensory neurons and revealing that discriminative choices must be based on bounded exact accumulation of evidence. We further demonstrate that this mechanism is not only necessary for the TIED to hold but is also sufficient to provide a virtually complete quantitative description of the behavior of the rats.

Main

In 1834, Weber reported that the just-noticeable-difference (JND) between the magnitudes of two stimuli is a fractional increment[1]. For instance, to identify with 0.75 probability the stronger of two pressure-stimuli applied to the skin, one must be 12% more intense than the other[2]. The fractional nature of the JND, termed Weber's law (WL) by Fechner[3], has been replicated in hundreds of studies across all sensory modalities and many animal species over the last two centuries[1,2,3,4,5]. The strong form of WL states that the probability of a correct discrimination between two stimuli—not just the JND—depends only on the ratio between their intensities.

WL embodies a non-trivial computation, since sensory receptors encode absolute stimulus magnitude explicitly in the form of monotonic increases in firing rate. How is the absolute magnitude of the stimulus factored out during discrimination? Although many explanations have been proposed[2,3,4,6,7,8,9,10], even the strong form of WL does not provide a sufficiently robust constraint to uniquely identify a particular mechanism. In our view, this insufficiency is because WL is a statement about a single aspect of discrimination: its accuracy. Surprisingly, although it was recognized early that reaction time (RT) is a key diagnostic in sensory discrimination[11], few studies have explored RT correlates of WL, and most models of WL have been cast within Signal Detection Theory[7] (SDT), which does not describe RTs. An exception is the work of Link[4], who proposed an explanation of WL in terms of bounded accumulation of sensory evidence,

a key concept in the modern study of perception and perceptual decision-making[12,13,14]. Although Link's model describes both accuracy and RT, it is only recently that the connection between WL and RT has started to be systematically explored[9,10]. Nevertheless, no clear empirical evidence currently exists that allows an unambiguous identification of the mechanistic foundation of WL.

Here, we report the existence of a psychophysical regularity—the TIED—which allows a tight quantitative specification of the effects of intensity ratios and absolute intensities on the process of sensory discrimination.

The TIED

We trained rats to discriminate the lateralization of binaural broadband (5–20 KHz) noise bursts, which were played through headphones to minimize uncontrolled stimulus variations (Methods and Supplementary Fig. 1a). Rodents use inter-aural level differences (ILDs) caused by the acoustic shadow of the head as the main binaural cue to localize sound in the horizontal plane[15]. ILD discrimination therefore engages a circuit for the comparison of sound intensities[16] that can be used to study WL[17]. We manipulated discrimination difficulty by varying ILDs pseudo-randomly across trials at a fixed average binaural level (ABL; Fig. 1c). To probe the effect of the overall sound level on ILD discrimination, ABLs were varied pseudo-randomly in blocks of 80 trials. We used a RT paradigm in which sound termination was triggered by the exit of the rat from a central port (Fig. 1a,b; Methods).

Because WL states that discrimination accuracy should only depend on intensity ratios, and since pressure root mean squared (RMS) ratios and ILDs (in dB) are equivalent (Fig. 1c), WL predicts that the accuracy of a given ILD discrimination should not depend on the ABL. Consistent with this prediction, psychometric functions at different ABLs overlapped (Fig. 2a,b). Differences in sensitivity (d' index[7]) for ILD discriminations at ABLs of 40 and 60 dB sound pressure levels (SPLs) did not reach statistical significance for any rat (all statistical comparisons of accuracy were based on Fisher's exact permutation tests, Bonferroni-corrected and evaluated at the 0.05

significance level, $n = 5$; Methods). For two out of the five rats, the d' index for ABL = 20 dB SPL was significantly different from that for 40 and 60 dB SPLs (in one case larger and in the other smaller; Supplementary Table 1). At the group level, differences in the d' values were again not statistically significant for any of the three ABLs. Thus, as shown previously for other species[17,18,19], rats display level-invariant ILD discrimination; that is, ILD discrimination in rats obeys WL.

Harder discriminations had longer RTs[12,13,14] (Fig. 2c), but discriminations of the same difficulty between sounds of lower ABL also took longer on average[9,10,18] (Fig. 2c). Both ILD and ABL had a significant impact on the mean RT on each individual rat (Supplementary Table 1) as well as at the group level (two-way repeated measures analysis of variance (RM-ANOVA); significant effect of ILD: $F(3,12) = 17.54$, $P = 0.0001$; of ABL: $F(2,8) = 77.12$, $P < 0.0001$). We found that changes in the ABL appeared to rigidly stretch the reaction-time distributions (RTDs) without changing their shape (Fig. 2d). Indeed, an appropriate uniform scaling of time (Methods and Supplementary Figs. 2 and 3) revealed that the RTDs are almost perfectly scale-invariant as a function of ABL for each difficulty and for all difficulties combined (Fig. 2e). In each case, more than 99% of the variance in the shape of the RTD for one ABL could be explained by a uniform stretching of the RTD for a different ABL (Fig. 2e, inset; mean $R^2 = 0.996$; Methods). Thus, for any given intensity ratio, discriminations appear to run faster when the absolute stimulus intensities are larger. This suggests the following general principle, which we refer to as time–intensity equivalence in discrimination or TIED: changes in the absolute intensity of two stimuli being discriminated under a fixed intensity ratio are completely equivalent to a change in the effective unit of time with which the discrimination duration is measured.

To probe the generality of the TIED, we first tested ILD discrimination accuracy in human subjects at two different ABLs (40 and 60 dB SPLs) with an effectively identical version of our rodent task. Of the 13 subjects tested, 9 were sufficiently unbiased and displayed level-invariant ILD discrimination (Methods) and were further analyzed. Human subjects were more variable than rats, and their RTs were more strongly affected by ILD than ABL. For the nine subjects with level-invariant accuracy (no

significant difference between the d' values across the two ABLs; Fisher's exact permutation test, $P = 0.98$), the TIED held well (Supplementary Fig. 4), although with slightly less precision than for rats. Still, RTDs at one ABL explained more than 99% of the variance in the shape of the RTDs for the other ABL (mean $R^2 = 0.992$). We also examined the RTDs from four rats performing an odor-mixture discrimination task, in which both the mixture contrast (ratio) and the overall concentration of the two odors were varied across trials[20]. This setup allowed comparisons of discrimination accuracy and RT for a given mixture ratio across different overall odor 'intensities'. We analyzed data from the conditions with the two highest concentrations, across which WL held (no significant difference between the d' values across the two concentrations; Fisher's exact permutation test, $P = 0.40$, $n = 4$). The data were again well described by the TIED (Supplementary Fig. 5; mean $R^2 = 0.994$). These results provide support for the generality of the TIED across two different species and sensory modalities.

The TIED specifies the mechanism underlying discriminations

The TIED is a stronger constraint than WL. Whereas WL effectively constrains the accuracy of an ILD discrimination at one ABL given its value at a different ABL, the TIED constrains not only the accuracy of this discrimination but also the full shape of its associated RTD. Thus, the TIED represents an opportunity to remove the ambiguity between different proposed explanations for WL. We pursued this approach by formalizing the TIED into a mathematical statement, and then searching for which mechanistic implementations of the discrimination process were compatible with it. To do this, the first step involved specifying a space of candidate models, ideally as large as possible. We considered the space of models in which choices are triggered when a decision variable (DV) hits for the first time either of two (possibly time-dependent) bounds—thus endowing the model with a well-defined RT—and where the DV evolves in time as a continuous Markov process (CMP)[21]. CMPs are very expressive and allow a variety of relationships between the DV and the sensory input, including close tracking of the instantaneous value of the evidence, attractor dynamics or perfect temporal integration. Most proposed models for perceptual decision-making are, or can be construed as, CMPs[4,8,22,23,24,25].

In a CMP, if one knows the value DV(t) of the DV at a given time t, then an instant later, the DV will be close to DV(t) and have a Gaussian distribution[21]. The mean and variance of this distribution are written as DV(t) + A(DV, t, s)dt and D(DV, t, s)dt, respectively, and can depend on DV(t), on time t and on the sensory input s, which confer the CMP framework its large flexibility. The quantities A and D are referred to as drift and diffusion coefficients of the process, respectively, and completely characterize the full temporal evolution of the DV. Our strategy to investigate the restrictions imposed by the TIED (see Supplementary Note) was to write down the expression for the drift and diffusion coefficients for two different CMPs associated with a discrimination between stimuli of magnitudes s_1 and s_2. The first CMP corresponds to a discrimination in which the overall intensity of the stimulus is changed while the intensity ratio is kept constant; that is, between stimuli of magnitudes ks_1 and ks_2. The second CMP corresponds to a discrimination between the original stimuli, but proceeding under a rescaled temporal variable $t' = \alpha t$. If the constant k is given and arbitrary, the form of all drift and diffusion coefficients such that the two CMPs are identical for some constant α quantitatively specifies the space of all CMPs consistent with the TIED. Mathematically, we seek functions A and D such that

$$A(\mathrm{DV}, t, ks_1, ks_2)dt = \alpha A(\mathrm{DV}, \alpha t, s_1, s_2)dt$$
$$D(\mathrm{DV}, t, ks_1, ks_2)dt = \alpha D(\mathrm{DV}, \alpha t, s_1, s_2)dt$$

Intuitively, equation (1) says that although the stimulus magnitude and the units of time have, in principle, different effects on the dynamics of the DV, the quantitative nature of these effects should be such that a constant scaling of the former can be perfectly emulated by a change in the latter.

We found that equation (1) is very restrictive, allowing only two types of solutions (Supplementary Note). The first type requires parameter fine-tuning and tight relationships between disparate, seemingly unrelated processes (for example, spiking statistics, rates of decay of the DV and time-varying decision bounds). While mathematically interesting, we deem this solution biologically implausible and do not consider it further. The only alternative solution requires the following four conditions: (1) a linear relationship between the variance and the mean of the sensory evidence, (2)

a power-law relationship between physical stimulus intensity and its internal representation, (3) a constant decision bound, and (4) perfect accumulation of evidence (that is, no intrinsic decay of the DV across time).

Condition (1) states that, effectively, the statistics of the sensory evidence should behave like a Poisson process. This is because the spike count statistics of a Poisson process are completely invariant if the rate and the count window are modified by multiplicative factors c and $1/c$, respectively. Thus, scaling the rate of a Poisson process can always be compensated by a rescaling of time. Of course, the TIED implies a scaling of stimulus magnitude, not firing rates, so an extra requirement is needed. This is provided by condition (2), as a power-law is the most general transformation between stimulus magnitude and firing rate such that multiplicative transformations in the physical intensity of sensory stimuli result in multiplicative changes in firing rate. The last two conditions effectively forbid any form of intrinsic dynamics of the DV or explicit time dependence in the discrimination process. As soon as these exist (barring the implausible solution we just discarded), it becomes impossible to find a rescaling of time that preserves the relative contribution of intrinsic and stimulus-dependent factors to the dynamics of the DV across all absolute stimulus magnitudes. For instance, if there is an intrinsic linear decay in the DV, as in leaky-accumulator models[23], the temporal rescaling factors that are necessary to compensate two different overall stimulus magnitudes will translate into two different effective intrinsic decay rates for the DV, thus destroying the scale invariance of the RTD across these two conditions.

We have tested that the same four conditions described above are sufficient for the TIED to hold in several of the most commonly used race models (two accumulators instead of one) in which the DV is higher dimensional[24].

The TIED predicts a breakdown of WL for short sounds

What do we exactly mean when we say that the effective unit of time of an ILD discrimination changes with the overall stimulus intensity? Larger sound intensities result in more activity in both sensory channels, leading to faster accumulation of sensory evidence toward the bound and faster RTs. However, as long as the four conditions described in the previous section are satisfied, the relative probabilities across time of hitting the bound at any particular moment become independent of overall intensity (Fig. 3a; Supplementary Note). This is formally equivalent to a change in the units of time of the discrimination process.

This mechanism makes the following counterintuitive qualitative prediction: if the DV reflects the temporal accumulation of evidence and if the DV is effectively evolving faster for louder sounds, then it should be the case that stopping the stimulus at times that are short compared with the typical RT will break the ABL invariance of performance, even at fixed intensity ratios. This is because, according to the TIED, the distributions of the DV at the offset of two short fixed-duration sounds of different ABLs should be the same as the distributions of the DV for two sounds of the same ABL but different durations (Fig. 3b). Since longer fixed-duration stimuli are expected to lead to higher discrimination accuracies under evidence accumulation[26,27], the TIED predicts that WL should break down for short fixed-duration sounds. Furthermore, this breakdown should become less pronounced as the fixed-duration stimuli become longer, because the likelihood that the DV will reach the bound approaches unity as the stimulus duration grows, and WL ensues.

We tested this prediction using low-intensity sounds (ABL = 20 and 40 dB SPLs) in a series of sessions in which the rats were still free to choose when to exit the central port, but the sounds terminated at a maximum sound duration (SD_{max}) if the rats were still at the central port by that time (Methods). Both aspects of the prediction were verified (Fig. 3c,d). Discrimination accuracy degraded for short SD_{max} for both ABLs ($n = 4$, two-way RM-ANOVA, significant effect of SD_{max}, $F(5,15) = 50.59$, $P < 0.0001$) but it degraded faster for the lower ABL (significant interaction between SD_{max} and ABL, $F(5,15) = 4.18$, $P = 0.014$; Supplementary Table 2), demonstrating that WL breaks

down in a sound-duration-dependent fashion (Fig. 3c,d). This result provides strong evidence to indicate that WL is the result of bounded accumulation of evidence. In fact, the data presented in Figs. 2 and 3 rule out an explanation of WL within SDT.

Sufficiency of the identified mechanism

How well is the behavior of the rats quantitatively specified by the mechanism inferred from the TIED? This is still an open question because only two qualitative properties of the behavior (level invariance of accuracy and scale invariance of RTDs) have been used to derive it. To address this issue, we constructed the simplest possible implementation of a CMP that obeys the four conditions identified above. Because the model is as simple as possible, this approach will not produce the best possible fits, but is instead intended to clarify the sufficiency of the mechanism to account for our observations at a quantitative level.

In our implementation (Fig. 4a; Supplementary Note), the evidence is the difference between the instantaneous activities of two sensory channels of N neurons. Neurons in each channel fire with Poisson statistics and have firing rates that grow as a power-law of the sound intensity at the corresponding ear. Accuracy and decision time in the model depend only on the following three parameters (Fig. 4a, equations (2) and (3)): the net gain (Nr_0) and the power-law exponent (λ) of the transformation between sound intensity and firing rate of each sensory channel, and the decision bound in units of the single-spike quantum of evidence (θ_e). To compare the model with the data, we also needed to specify a stimulus-independent non-decision time (t_{NDT}), which accounts for sensory and motor delays that contribute to the measured RTs. This is the absolute minimum number of parameters necessary to fit a bounded accumulation model with nonlinear stimulus encoding. Close to psychophysical threshold, the choice dynamics in this model are captured by a single-parameter diffusion equation (Supplementary Note) as follows:

$$\frac{dz}{d\tau} = \Gamma \mathrm{ILD} + \eta(\tau)\theta_z = \pm 1; z(\tau = 0) = 0$$

where z is a normalized DV (in units of θ_e), and the single parameter $\Gamma = \lambda\theta_e/(40/\log(10))$ specifies the map between the stimulus ILD and the strength of evidence. The ABL of the stimuli does not appear in equation (2), which implies that all the non-trivial properties of the discrimination depend exclusively on the intensity ratio. The temporal variable τ in equation (2) measures the duration of the discrimination in an ABL-dependent effective unit of time t_θ (ABL). Therefore, $\tau = t/t_\theta$ (ABL) with

$$t_\theta(\mathrm{ABL}) = \frac{\theta_e^2}{2Nr_0} 10^{-\lambda \mathrm{ABL}/20}$$

The prediction of the model for the RTs is therefore $\mathrm{RT} = t_{\mathrm{NDT}} + \tau_{\mathrm{DT}} t_\theta(\mathrm{ABL})$, where τ_{DT} is the stochastic decision time associated with equation (2). Equations (2) and (3) specify the effect of difficulty and absolute intensity on sensory discrimination (see also Supplementary Fig. 6). The equivalence between time and absolute intensity, which defines the TIED, is explicitly captured by equation (3).

In addition to having the least possible number of parameters, the model is easily falsifiable, since the psychometric function and the shape of the RTDs for each difficulty all depend on the single parameter Γ. We specified the value of Γ from fits to the discrimination accuracy of the rats. The single-parameter fit accounted for 99.2% of the variance in the psychometric data of the rats (Fig. 4b). The JND of our rats was ~2.2 dB (Fig. 4b; Supplementary Table 3), which is similar to that measured in several other species[28,29]. At this point, the shape of the RTDs and their dependency on difficulty are fully specified, and can therefore be seen as model predictions. The only remaining degrees of freedom in our fit are t_{NDT}, which shifts all the RTDs by the same amount, and t_θ(ABL), which can rigidly stretch the RTDs for each ABL separately. We used the empirical RTDs to infer the parameters specifying t_{NDT} and t_θ(ABL) (Methods), but we again challenged the predictive power of the model by using only data from the ABL = 20 and 60 dB SPL conditions. Still, the full shape of the RTDs for all conditions—including those for ABL = 40 dB SPL, which are a pure model prediction—are accurately captured by the model (Fig. 4c; see Supplementary Fig. 7 for

fits of single-rat data), with model RTDs accounting for $97.7 \pm 2.3\%$ of the variance of the observed RTDs (mean $\pm$ s.d. across the 12 conditions; Methods). The accuracy of these predictions on data that were not used for the fits suggests that the relationship between RT and accuracy on the one hand and RT and intensity on the other is of the same nature in the model and in the experiments.

The fitted value of the power-law exponent $\lambda = 0.089 \pm 0.004$ (see Supplementary Table 3 for all parameter values, which were robust using various fitting methods; Supplementary Figs. 8 and 9), indicates a large degree of compression, which results in a relatively mild dependence of the RTs with ABL. The value of $\theta_e = 48.3 \pm 2.3$ implies that each sensory spike contributes a few percent of the evidence necessary to reach threshold. Finally, the value of Nr_0 implies a range of firing rates of ~4 to 7 KHz on each sensory channel across the stimuli in our experiment.

Further controls and analyses demonstrated that the behavior of the rats is consistent with our implicit modeling assumptions. Rats could discriminate the ILD of 10 KHz pure tones at 60 dB SPL and broadband noise with the same accuracy (Fig. 5a), which shows that they exclusively use the ILD of the sound to perform the task. The discrimination accuracy for 'frozen noise' bursts was identical to control conditions (Fig. 5b), thus showing that performance-limiting noise has an internal origin. Unexpectedly, the behavior of the rats was virtually independent of the trial history, establishing that choices on a given trial depend almost exclusively on the nature of the sensory stimulus (Fig. 5c; Supplementary Fig. 10). These findings confirm that the ILD discrimination task we are using is an appropriate paradigm to read out the percepts of subjects from their behavioral choices with minimal confounding factors.

The Weber fraction in ILD discrimination

The Weber fraction (WF) is typically used to quantify the minimal discriminable stimulus increment at the psychophysical threshold; that is, when accuracy is not limited by motivational factors. To investigate what determines the WF in ILD discrimination, we performed a series of behavioral manipulations with the goals of establishing whether our rats are indeed at the psychophysical threshold and of clarifying other potential factors, beyond motivation, that might be limiting choice accuracy in our task. First, we attempted to increase the discrimination accuracy by increasing the reward magnitude for correct discriminations in the hardest conditions relative to the easiest ones (Methods), but this failed to induce changes in performance (Fig. 5d, comparison of d' values between uneven reward and control conditions: $P = 0.9$, Fisher's exact test, $n = 4$). Next, we presented blocks with only the two most difficult conditions, a manipulation expected to improve discriminability due to a motivation-enhancing decrease in reward rate, a narrower prior on stimulus difficulty[25] and a potential removal of uncertainty about the location of the categorization boundary[20]. Nevertheless, accuracy again did not change (Fig. 5e, comparison of d' values between only hard and control conditions: $P = 0.5$, Fisher's exact test, $n = 3$). To confirm this result, we performed longer-lasting (more sessions) manipulations of stimulus difficulty, and we also presented only easy (in addition to only hard) blocks of stimuli (inducing changes in experienced reward rate of more than 100%; Fig. 5f, inset), but we observed the same negative results (Fig. 5f; see Supplementary Table 2 for statistics for all behavioral manipulations). These findings show that accuracy in our task is not significantly limited by the uncertainty of the rats about the indifference point or by stimulus priors. They also show that accuracy is not limited by motivation, which indicates that rats are performing the task at the psychophysical threshold.

Can the TIED, together with inferences drawn from these behavioral manipulations, be used to shed light on the mechanism that sets the WF in our task? At a mechanistic level, the precision of a decision maker trying to maximize accuracy at the expense of RT can be limited in the following three qualitatively different ways: through the accumulation mechanism (that is, in the presence of leak), by the statistics of the evidence (that is, in the presence of temporal correlations) or by the magnitude of the

decision bound (as accuracy increases with the magnitude of the bound). Which of these three mechanisms are at play in our task?

Both leak and a fixed $\tau_{\text{corr}} \sim \text{RT}$ are incompatible with the temporal scale-invariance required for the TIED and are also not necessary to provide a satisfactory account of the choice and RT data (Fig. 4b,c), which suggests that they do not play a large role in setting the WF in our task. Across-trial variability in the evidence can result from trial-to-trial updating of the categorization boundary[20], but the $\text{ILD} = 0$ boundary is hardwired, and we found no evidence of updating (Fig. 5c,e,f; Supplementary Fig. 10), again suggesting that this process is not a critical source of uncertainty in our task. Slow fluctuations in the excitability, or baseline firing rate, of the sensory neurons encoding the stimulus will also produce across-trial variability in the evidence. Multiplicative (but not additive) noise in baseline firing rates is consistent with the TIED (Supplementary Note), so we considered its effect quantitatively, despite the fact that it also did not seem necessary for a good fit (Fig. 4b,c). We modeled the firing rates in the two channels in trial i as $r^i_{\text{L,R}} = r_{\text{L,R}}(\text{ABL}, \text{ILD})(1 + \sigma_r n^i_{\text{L,R}})$, where $r_{\text{L,R}}(\text{ABL,ILD})$ are the sensory firing rates in the absence of baseline variability, niL,RnL,Ri are normal random variables and σ_{r} measures the magnitude of the multiplicative noise (Methods). Adding trial-to-trial variability only marginally affected the value of the model parameters. The best fit value of σ_{r} was 0.0063 ± 0.0091 (95% bootstrap confidence intervals (CI)). The decision bound changed from 45.7 ± 2.0 to 46.7 ± 4.1 (both 95% bootstrap CI). If we artificially set σ_{r} to zero and left all other parameters at their best-fit values, the JND decreased by just 1.7%. If, on the other hand, we doubled the magnitude of the bound, the JND decreased by 47%. We therefore conclude that trial-to-trial variability in baseline rates makes, at most, a marginal contribution to the WF in our data. Having ruled out a significant contribution from inefficiencies in the process of accumulation or from non-averageable sources of variability in the evidence, the alternative is that the WF is simply set by an upper limit of the evidence bound. Such a limit is consistent with the apparent inability of the rats to improve discriminability in response to changes in reward magnitude (Fig. 5d) or reward rate (Fig. 5e,f).

Discussion

We have shown that while, in agreement with WL, changes in the absolute magnitude of stimuli being discriminated at fixed intensity ratios have no effect on discrimination accuracy, they have a clear and quantitatively precise effect on RT, leading to an almost-perfect rescaling of the RTD (Fig. 2a,d,e; Supplementary Figs. 3–5). Thus, changes in absolute stimulus intensity have the same effect as changes in the units of time within a trial, a phenomenon we have referred to as TIED. The TIED is consistent with the well-known fact that more intense sensory stimuli are perceived as lasting longer[30].

Unlike WL, the TIED is sufficiently restrictive to narrowly specify the computational mechanism underlying ILD discrimination, which is characterized by bounded exact temporal accumulation of evidence, Poisson-like variability and power-law encoding of stimulus intensity. This mechanism is not only necessary to account for the TIED: in its simplest possible implementation it is also sufficient to account for 98–99% of the variance in the accuracy and RT of the rats (Fig. 4b,c; Supplementary Fig. 7). To interpret this number, note that there are many model modifications that would have quantitatively improved the accuracy of our fits at the cost of more parameters—such as trial-to-trial variability in the bound or in the baseline sensory rates (Supplementary Note)—or of small quantitative violations of the TIED—such as trial-to-trial variability in non-decision time. We show that only a very small fraction of the variance in the full choice and RT data remains to be accounted for by elaborations of the simplest mechanism consistent with the TIED.

Effect of overall stimulus intensity in sensory discrimination

Overall sound level can be considered a ‘nuisance parameter’ for the identification of the location of a sound source. Indeed, ABL changes the gain of ILD on the average sensory evidence (equation (40), Supplementary Note). Gain normalization[31] is often used to remove the nonspecific effect of overall stimulus intensity, including in ILD discrimination[28]. Interestingly, the mechanism we studied can be interpreted as a form of gain control that does not require explicit normalization. Under Poisson variability (observed in the auditory localization pathway[32,33]), although drift rates increase with

ABL, so does the noise in the trajectories, resulting in exactly the same probabilities of hitting either bound[4]. This ensures level-invariant accuracy (Fig. 3a) without the need of any explicit normalization mechanism.

Studies of the neural basis of level-invariant ILD discrimination have focused on finding explicit level-invariant codes (for example, neurons with the same ILD tuning regardless of ABL). Neurons in the lateral superior olive, the inferior colliculus or the auditory cortex under anesthesia do not tend to be explicitly level-invariant[34,35,36], although opponent-process models (which consider linear decoders subtracting the activity of neurons with opposite ILD tuning) achieve higher levels of invariance[34,37,38]. Our results demonstrate that certain important behavioral invariances do not need to be explicitly represented in the brain: the TIED is a purely computational property based on a threshold-crossing mechanism, and no static decoder of the sensory neurons or of the DV in our model shows the same qualitative properties as the behavior.

WL and RT

This fact also has implications concerning the feasibility of explanations of WL based on SDT. Most previous explanations of WL can, either explicitly or implicitly (many explanations of WL preceded in time the development of SDT), be cast within SDT, in the sense that they do not view the decision of when to stop sampling the stimulus as critical for the explanation of WL. Our results present serious challenges to any type of explanation of this sort. Since the sampling duration in SDT is an external parameter outside the decision process, these models can simply provide no explanation for any RT phenomenology (Fig. 2). The breakdown of WL for short fixed-duration stimuli but not RT choices (Fig. 3) also seems fundamentally at odds with any SDT-type explanation. In general, the lack of dynamics in SDT implies that the invariances of the behavior need to be explicitly encoded in the form of the DV in models within this framework.

In contrast to SDT, Sequential Sampling Theory (SST) explicitly addresses the problem of when to stop sampling the evidence in a categorical choice[12]. Link[4] was the first to emphasize a connection between SST and WL, highlighting the importance of Poisson

variability. More recently, some studies[9,10] have addressed the relevance of RT as a key diagnostic for understanding the effect of absolute intensity in sensory discrimination. Simen and colleagues[10] realized that Link's model predicts that WL should be associated with scale invariance of the RTD, thus highlighting the connection with temporal production[39,40]. Empirical evidence was presented that, although quite variable, was generally supportive of the qualitative claims. Simen and colleagues did not attempt a quantitative match between their model and data. Our work extends the contributions of Link[4] and Simen et al.[10] in important ways. First, we provided necessary conditions for the TIED to hold, clarifying the space of models it is compatible with. Second, the precision of our main experimental finding (Fig. 2; Supplementary Figs. 3–5) allowed us to demonstrate the quantitative sufficiency of the identified mechanism to provide an almost complete description of the behavior (Fig. 4). Less precise experiments would have required extra assumptions beyond the TIED to accurately fit the data. Third, we validated the counterintuitive breakdown of WL for short fixed-duration sounds (Fig. 3). Finally, the strict constraints imposed by the TIED together with our behavioral manipulations (Fig. 5) allowed us to delineate the mechanism that sets the WF in our task.

We found that a key computation in sensory discrimination is a 'perfect' (that is, without time decay) accumulation of evidence, consistent with work from Brody and colleagues[27] in which a task was explicitly designed to probe the temporal accumulation of discrete sensory pulses. They have recently addressed the relevance of absolute stimulus intensity for discrimination, demonstrating that accuracy degrades with the total number of pulses being discriminated when the difference between left and right is fixed, consistent with WL[41]. Although these results could be explained with a SDT model that required using multiplicative-noise statistics, which is consistent with long temporal correlations in the evidence, our results and those of Simen[10,40] show that the resulting scale invariance of the behavior can be obtained with temporally uncorrelated evidence samples as long as bounded accumulation is used to trigger choices. At present, a quantitative comparison between our results and those of Brody and colleagues is difficult given significant differences between the two tasks, especially at the level of the behavioral report (choice at different fixed durations versus choice and

RT) and at the level of the target for inference ('latent' ILD versus explicit difference in the number of pulses).

Several studies have suggested that leaky evidence accumulation[23] provides a better account of the data in some sensory discrimination tasks. Although the TIED is incompatible with leaky accumulation (Supplementary Note), our results do not necessarily conflict with these studies, as leaky accumulation, unlike WL, tends to work best in situations in which the evidence is non-stationary[23,42].

Scale-invariant distributions and power-law transformations

Scale invariance of response distributions is a robust feature of temporal production tasks[39]. The accumulation of neural activity with Poisson-like statistics up to a single bound has also been shown to generate scale-invariant distributions in this context[10,40], although the nature of the accumulated signal in temporal production, in contrast to sensory discrimination, is not yet completely clear. Scale invariance of RTDs with respect to changes in difficulty (as opposed to absolute intensity) has also been previously considered[43]. However, unlike the case of absolute intensity, this rescaling is only approximate (Supplementary Figs. 3 and 11).

Scale invariance as captured by the TIED requires a power-law transformation between stimulus intensity and sensory firing rates (Supplementary Note). Power-law transfer functions have often been reported in many sensory receptors[44,45] and are believed to be important for compressing the large dynamic range of sensory signals. Power-laws also have a privileged role in psychophysics thanks to Stevens[5,46], who proposed that the subjective magnitude of a stimulus has a power-law relationship with physical stimulus intensity. It is intriguing that our results suggest that physical sensory intensity is also encoded as a power-law. Studies of WL, starting with Fechner, have also sometimes invoked a logarithmic transfer function[3,47]. In our setting, such a model predicts that there should be no changes in RT as a function of ABL, and is thus not compatible with our findings. Logarithmic transformations play no particular role in our results (Supplementary Fig. 6).

Specificity of results to ILD discrimination

We believe that the accuracy with which the experiments specified the mechanism driving the behavior depends on the ethological significance of the task. In ILD discrimination with respect to the midline, we are recruiting a circuit designed by evolution for the purpose of comparing stimulus intensity, rather than trying to create or co-opt a general purpose comparison mechanism with no particular significance for the rat. Furthermore, the relative weight of the left and right sensory channels, which specifies the decision variable, is hardwired[16], which probably underlies the lack of an effect in our manipulations of stimulus difficulty (Fig. 5e,f), as the rat brain has an accurate representation of the auditory midline independently of our task. In contrast, arbitrary boundaries (such as a particular odor mixture[48]) need to be learned by trial and error through reinforcement[20], which introduces an additional source of uncertainty ('decision rule' uncertainty) in discrimination. We hypothesize that the lack of decision-rule uncertainty, together with the similarity between the task contingency and a hardwired behavior (orient toward the source of a sound) and the lack of rules enforcing response timing, all contribute importantly to the ability of the rats to generalize (Fig. 5a) and to display negligible lapse rates (Fig. 5b,f) or sequential dependencies in performance (Fig. 5c; Supplementary Fig. 10). The addition of decision-rule uncertainty for a given amount of sensory uncertainty (for example, ILD discriminations with respect to a non-zero boundary) is expected to also directly impair discrimination accuracy, as has been previously observed[49].

The limits of discrimination accuracy

Understanding the mechanisms limiting the accuracy of sensory discriminations using psychophysics has been a difficult task due to the diversity of potentially contributing factors[2,4,50]. Our results (Figs. 2, 4 and 5) suggest that discrimination accuracy in our task is effectively limited by the value of the evidence bound. This finding has a number of salient implications. First, since the bound is not adaptive in the face of manipulations, it cannot be optimal in the context of imposed task contingencies (although its value may still be adaptive in the face of longer-lasting developmental or evolutionary constraints). Second, the control that motivation exerts over the value of the bound appears to have some fundamental limits, and is thus not universal, as

generally assumed in SST. Finally, nothing in our experiments provides insight into what is special about the particular value of the decision bound that we found. How the particular value of the bound is set and under which conditions motivation loses control over its value remain open questions to be elucidated in future work.

Methods

Experimental animals

All procedures were reviewed and approved by the animal welfare committee of the Champalimaud Centre for the Unknown and approved by the Portuguese Direcção Geral de Veterinária (reference no. 0421/000/000/2019). Experiments were performed on 11 adult female Long–Evans hooded rats. Animals were 12–13 weeks old and weighed between 250 and 300 g at the beginning of the experiments. They were kept above 85% of the initial weight throughout. All animals were naive to any behavioral tests. Rats had free access to food, but water was restricted to the behavioral sessions, which were conducted during five consecutive days per week. Animals had access to water during the sixth day and were deprived of water for 24 h before each round of five sessions. No statistical methods were used to predetermine the number of animals or the number of trials per animal, but our sample sizes were similar to those reported in previous publications[48]. We used a 'within subject' design in which the animals were tested in all experimental conditions, so there was no need to apply blinding or randomization. All results in the main text came from the same batch of rats (Batch A; see Supplementary Table 2) except for the data presented in Figs. 3 and 5b,f, which came from a second batch of animals (Batch B). The five rats in Batch A were tested in the RT sound lateralization task. Four of them performed blocks that included only the hardest conditions, and three of them performed blocks with uneven reward and blocks with pure tones. The six rats in Batch B were tested in the bidirectional motivation manipulations and in the frozen-noise manipulations. Four of them were tested in the capped sound duration sessions.

Auditory stimuli

A percept of lateralization was created by presenting cosine-ramped (10 ms) broadband (5–20 kHz) noise bursts with different intensities to each ear (ILD). The noise was independently generated for each ear and for each presentation (except for the results shown in Fig. 5b) using a Tucker-Davies Technologies RP2 module at a sample rate of 50 kHz. Headphones were calibrated weekly, using a Brüel & Kjaer Free-field one-quarter-inch microphone placed in front of the speaker, 5-mm apart.

Behavioral apparatus and headphone design

Rats were trained and tested on the sound lateralization task (Fig. 1a,b) using a standard Coulbourn Instruments modular box (30 × 25 × 30 cm) with three ports (8-cm apart). All components of the behavioral setup were connected to a RP2 module and accessed by a computer running Matlab 2012b (https://www.mathworks.com) using TDevAcc controls. The behavioral setup was placed inside a soundproof box that was illuminated by infrared lights and equipped with an infrared camera to observe the animals during the sessions.

Both the base to be implanted in the skull and the structure of the headphones were designed using Sketchup (https://www.sketchup.com) and 3D-printed in VisiJet(R) EX200 Plastic. The speakers (Knowles, no. 2403 260 00029) were aligned with the ears and placed 5 mm from the entrance of the ear canal. Once adjusted, all pieces were glued together and remained fixed throughout the experiment (Supplementary Fig. 1a). The headphones were attached to the base at the beginning of each behavioral session and detached before taking the animal back to the holding cage.

Behavioral tasks

Sound lateralization task: temporal and outcome contingencies

Rats started a trial by poking in the central port within a 6 s time window (start trial waiting time) triggered by the end of the inter-trial interval (3 s), which was signaled by a light in the box turning off. After a short, variable fixation time (FT, 300–350 ms) the sound was played binaurally, through the headphones, until the rat left the central port or until the maximum presentation time (6 s) was reached. Rats had to communicate,

within a 2 s time window (response waiting time), whether the sound was louder at the left or right ear by poking with the snout in either the left or right ports, respectively. Correct choices were rewarded with a drop of water (28 μl) and incorrect responses were penalized with a 10-s timeout, during which the rat was not able to start a new trial. Trials in which the rat failed to start a new trial within the start trial waiting time, broke fixation during the FT or failed to poke in either lateral port within the response waiting time were considered aborts. Aborts were repeated after a 1-s time penalty.

Each session was divided into blocks of 80 trials. Within each block, the ABL was kept constant, while the ILD changed pseudo-randomly from trial to trial. Typically, sessions lasted for 2 h and rats performed between 800 and 1,200 trials.

Sound lateralization task: training

Animals were initially trained in a simplified version of the task, in which fully lateralized sounds (50 dB SPL broadband noise) were presented from either of the arena speakers. Rats quickly understand the basic contingency of the task, within a few hundred trials. The sound was played until the animal entered one of the lateral ports, and errors were repeated immediately. Short fixation times and long waiting times were used to increase the chances of the rat to complete the trial while exploring the box. Every time the rat completed a trial, the fixation time was increased by 1 ms and, once the rat completed three consecutive blocks (120 trials per block) with less than 30% abort rate, waiting times were set to their final durations and ILDs were introduced. Initially, the ABL was set to 50 dB and the ILD step was set to 4 dB; depending on the performance of the animals, the step was decreased gradually until the final 1.5 dB. Once performance with ABL of 50 dB and ILD step of 1.5 dB was stable, the magnetic base for the headphones was implanted and the animals were allowed to recover for at least 1 week, during which they had free access to food and water.

Block types used for the different tasks

To test various hypotheses about the nature of the behavior, we modified the above-described basic task in several ways. Rats from Batch A were tested in the following five types of blocks: A1 'standard' blocks, in which four ILDs (of each sign) linearly spaced from 1.5 to 6 dB steps were presented. All data in the main text except for those presented in Figs. 3 and 5a,b,d,f came from A1 blocks. A2 'hard' blocks, in which only the four ILDs closer to the midline (±1.5 and ±3 dB) were presented (Fig. 5e,f). We did not only present the hardest condition because when we attempted this, the rats 'gave up' and became biased. A3 'uneven reward', in which we increased (decreased) the amount of water delivered after correct discriminations for the two hardest (easiest) ILD conditions by 20% (Fig. 5d). A4 'log noise', in which five ILDs (of each sign), logarithmically spaced between 1 and 8 dB, were used (Fig. 5a). A5 'log pure tones', identical to the previous block, but using 10 kHz tones instead of broadband noise (Fig. 5a). In all types of blocks, except for A5, three ABLs (20, 40 and 60 dB SPL) were used. For A5, only ABL = 60 dB SPL was used. Rats from Batch B were tested in the following five types of blocks with ABL = 50 dB SPL: B1 'standard', in which four ILDs (of each sign) logarithmically spaced between 1 and 8 dB were used (Fig. 5b,f). B2 'hard', in which only the two most difficult conditions (1 and 2 dB) were used (Fig. 5e). B3 'frozen noise coherent', in which the exact broadband sound (out of four different examples) was presented in the two headphones appropriately scaled to produce a given ILD (ILDs were the same as in the standard blocks). B4 'frozen noise non-coherent', in which a different one of the four examples was played in each headphone (same ILDs). Blocks B3 and B4 where used in Fig. 5b. B5 'easy sessions', in which only ILDs = ±4.5, 6, 9 and 15 dB were used (Fig. 5f). After these manipulations, four remaining rats from this batch were initially trained in a variant of the standard RT task with a different set of conditions (ILD = 0, 0.5, 1.25, 2.25, 4 and 8 dB and ABL = 10, 25, 40, 55 and 70 dB SPL) designed to sample the effect of difficulty and intensity more densely and without blocks. After a few sessions, we decided to perform the capped sound duration experiment and we then switched conditions to ILD = 1, 2, 4 and 8 dB (of each sign) and ABL = 20 and 40 dB SPL. The range of SD_{max} tested was 50, 75, 100, 150 and 250 ms. These sessions are used in the results presented in Fig. 3.

Task variants: discrimination of the ILD of pure tones

Three of the rats were tested with the A4 and A5 blocks mixed pseudo-randomly.

Task variants: manipulations of motivation

For rats in Batch A, A2 blocks were randomly included among standard blocks within a set of sessions in such a way that every time hard conditions were tested, at least three consecutive blocks (one for each ABL) were used. Results from this manipulation are shown in Fig. 5e. These same rats also performed a series of sessions in which all blocks were the A3 type (Fig. 5d). Longer-lasting and bidirectional manipulations of motivation were tested with rats of Batch B. These rats were tested in a series of sessions in which an initial standard block was followed by the B2 blocks until the end of the session (Fig. 5f).

Task variants: external versus internal noise

We selected four broadband noise samples and either used the same or different samples at the two ears. Rats of Batch B were tested in a series of sessions with the B3 and B4 blocks alternating pseudo-randomly within the same session (Fig. 5b). Note that even with frozen-noise stimuli, waveforms at the two ears will not be perfectly coherent because the headphones are not inserted in the ear canal.

Task variants: maximum sound duration

As shown in Fig. 3, the task was still in RT configuration but, for a given SD_{max}, the sound was stopped at that SD_{max} if the rat had not left the central port at that time. For choices with $RT < SD_{max}$, the sound offset was triggered by the central port exit. Rats performed the task in mini blocks of 16 trials. Each mini block contained two permutations of all the ILDs at fixed ABL and SD_{max} values, and ABL and SD_{max} values were chosen randomly from their possible values for each mini block.

ILD discrimination task for human subjects

All procedures were reviewed and approved by the Ethics Commission for human studies of the Champalimaud Centre for the Unknown. Human volunteers performed an ILD discrimination task at two different ABLs. The experiment took place in a small, closed room. A computer running Matlab controlled the stimulus delivery through custom-written software using Psychtoolbox-3 (http://www.psychtoolbox.org). The auditory stimuli were delivered through a pair of over-the-ear headphones (Sennheiser HD 300 PRO) that were regularly calibrated. Subjects interacted with the behavioral software using a bimanual gamepad (Logitech Gamepad F310).

Sounds were cosine ramped (10 ms) wideband noise bursts (1–15 KHz) independently generated for each ear. We used ILDs of 0.3, 0.6, 1.2 and 2.4 dB, and ABLs of 40 and 60 dB SPLs. As in the rat experiment, ILDs varied pseudo-randomly across trials and ABLs varied in blocks (duration of 120 trials). Subjects performed 240 auditory trials per session, which correspond to two blocks of different ABLs.

Each trial began when the subject pressed both shoulder buttons (or bumpers) on the gamepad to indicate readiness. The stimulus was presented after a variable fixation period (of duration equal to the sum of 700 ms and an exponentially distributed (mean 300 ms) duration truncated at 1.5 s) and played for as long as the subjects kept the shoulder buttons pressed. Responses were signaled by releasing the button on the side on which the sound was louder. Trials with RTs shorter than 100 ms or longer than 3 s were discarded. Subjects received feedback after each trial. After a correct response, the subsequent trial was ready after a short inter-trial interval of 0.5 s; incorrect responses were followed by a short, 3.5-s timeout. Subjects received feedback about their mean accuracy over the session with a numerical display that was present on a corner of the screen.

After a few initial training sessions, subjects maintained above 80% accuracy on most sessions, and only these sessions after performance and RTs were stable were used for the analysis. Subjects completed an average of nine sessions in steady-state conditions. Payment for the sessions was not linked to performance, and all participants were compensated with 2.5 euros per session.

We collected data from 13 healthy subjects with normal hearing, 5 of them female, aged between 20 and 40 years. Four of the subjects were excluded based on the following criteria. Two subjects were strongly biased (one of them more than three across-subject standard deviations away from the across-subject mean in criterion, the other more than three across-subject standard deviations away from the across-subject mean in the relative difference in average RT between left and right choices). Two other subjects clearly violated WL (relative difference in discriminability $\Delta d'/\text{Mean}(d')$ larger than 0.25).

Olfactory discrimination task

Details of this behavioral task have been previously published[20]. We analyzed their data for the conditions at the two highest concentrations (10^{-2} and 10^{-1} (v/v)) for which WL held. We did not perform any subsequent data selection or pre-processing beyond what was described in that study[20].

Statistics

To determine the minimum RT to be analyzed, we used a two-sample Kolmogorov–Smirnoff test. For assessing the ABL dependence of ILD discrimination accuracy, we used Fisher's exact test, two-tailed, Bonferroni-corrected. Two-way ANOVA tests (3×4) were used to describe the effects of ABL and ILD on RTs. Group-level comparisons were made using two-way RM-ANOVA tests. For the limited sound duration experiments, we used two-way RM-ANOVA tests (2×6) to describe the effects of ABL and SD_{max}. We also used a two-tailed *t*-test, Bonferroni-corrected, to asses the significance of the difference in performance across the two ABLs for each SD_{max}. To evaluate the accuracy of the scale invariance of the RTDs as a function of ABL and ILD, we used two-tailed Fisher's exact test. We tested normality and

homoscedasticity for the ANOVA tests shown in Fig. 3 (Lilliefors test for normality, $P = 0.1$, $n = 4$; Levene's test for homoscedasticity, $P = 0.56$, $n = 4$). We did not test normality for the RT ANOVA tests (Fig. 2) because we have an actual accurate generative model of RT. Specific details on each testing procedure are given in the corresponding part of the data analysis section below.

Data analysis

Isolating stimulus-dependent RTs

To exclude those trials in which behavior was not driven by the stimulus, we looked for the minimum RT for which there was evidence of condition dependence. To this end, we used two-sample Kolmogorov–Smirnoff tests to compare the distribution of RTs corresponding to the two conditions with the shortest (ILD = 6 dB, ABL = 60 dB SPL) and the longest (ILD = 1.5 dB, ABL = 20 dB SPL) mean RT. Starting at 50 ms, we systematically included longer and longer RTs. As evident in Supplementary Fig. 2a,b, if the maximum RT is sufficiently short, the two distributions are not significantly different but for RTs > RT_{min} = 90 ms, they become different. For all analyses, we excluded trials with RT < RT_{min}. In addition, since the shape of the RT distributions is very well behaved and understood in our study, we also excluded trials that had exceedingly large RTs, which presumably reflect disengagement. For all analyses except model fitting, we chose a conservative value of RT_{max} = 1,000 ms (the fraction of trials with RT > 1,000 ms was always very small: $0.39 \pm 0.39\%$, mean ± s.d., across rats). For model fitting, trials with RTs above the 97% percentile in the RT distribution of each rat were excluded. Empirical estimates of the RT distribution (Figs. 2 and 4; Supplementary Fig. 7) were made using kernel density estimation[51] as implemented with a custom Matlab code.

Accuracy

Accuracy was assessed using the SDT statistics for sensitivity (d') and criterion (c)[7] within each 80 trial block. Level dependence was tested by comparing d' and c as a function of ABL (three comparisons: 20 versus 40 dB; 20 versus 60 dB; and 40 versus 60 dB) using Fisher's exact test with Bonferroni correction. Because easy blocks barely contain any errors, for comparisons between easy and standard blocks (Fig. 5f), the

statistic was percent correct instead of d', and significance was evaluated again using Fisher's exact test.

As shown in Fig. 3, we fit a two-parameter logistic function (slope and asymptote; for fixed-duration sounds, the rats sometimes displayed non-negligible lapse rates) of the across-rat mean percent correct as a function of difficulty. We used two-way ANOVA to quantify the contribution of sound duration and ABL to accuracy in these experiments, using d' as the statistic. Since the RT sessions with multiple ABLs in this batch used different values of ILD compared to the sessions with capped sound duration (see above), we fit a four-parameter logistic function (two asymptotes, slope and bias) to the behavior of each rat on the RT sessions. We then estimated by interpolation the performance of each rat at the same values of ILD used in the SD_{max} sessions and created datasets with the same number of trials as the actual RT sessions, but with ratios of correct and incorrect trials consistent with the interpolated performances. These new datasets were then used for calculating the d' values of the RT sessions. Comparisons were made using two-way RM-ANOVA.

Accuracy of temporal rescaling

We assessed how accurately the RTDs for different experimental conditions resembled a uniform scaling of time (Fig. 2d,e; Supplementary Fig. 3) using the fact that when two distributions are related by a uniform scaling, their quantiles are proportional (Supplementary Fig. 3). For fixed ILD (ABL), we regressed the percentiles of each condition on those of the fastest, that is, 60 dB SPL (6 dB). Since the RTDs are significantly right-skewed, higher percentiles have more uncertainty. Thus, we used weighted least-squares to perform the linear fit. To compute the uncertainty in the percentiles, we generated 1,000 bootstrap re-samples from the RTD of each rat for each condition. For each re-sample, we computed the percentiles and averaged them across rats, obtaining a sampling distribution of 1,000 percentiles. These distributions are very well approximated by Gaussians. To avoid the complication of having uncertainty in both the dependent and independent variables, we followed the following heuristic approach: we assumed that the independent variable (percentiles of the fastest condition) had no uncertainty, and assigned an uncertainty $\sigma_D^{\text{eff}} = \sqrt{\sigma_I^2 + \sigma_D^2}$ to the

dependent variable, where $\sigma^2_{I,D}$ represents the variances of the sampling distribution of the percentiles of the independent and dependent variables, respectively. We then applied the standard weighted least-squares algorithm to find the value of the slope and its associated R^2, which we report in Fig. 2e, inset, Supplementary Fig. 3. The shaded areas in Supplementary Fig. 3a,b correspond to $3\sigma_D^{\text{eff}}$.

History effects

We used logistic regression[52] to investigate the effect of trial history on the performance of the rats in single trials from the standard A1 blocks. We used the model to predict the performance of completed trials within the valid range of RTs. We used 12 predictors from the current trial, which corresponded to the four conditions of difficulty (absolute value of ILD) at each ABL (each coded as a [−1, 0, 1] identifier, with 0 representing that a different ILD–ABL pair was presented in that particular trial and ±1 representing a positive (negative) value of ILD). The following four aspects of trial history were evaluated: history of the stimulus lateralization (the sign of ILD regardless of ABL), the history of response side, and two response–outcome predictors (history of response side after correct choices, and after errors). For a given trial *i*, the stimulus history predictor for trial $i-j$ was coded as a [−1, 1] indicator, where 1 (−1) indicates that the stimulus at trial $i-j$ had positive (negative) ILD (by convention, we predict the choose-right probability 'R', that is, the probability of making a response to the side that is correct when ILD is positive). The response history predictor was coded as a [1, 0, −1] indicator, where 1 (−1) indicates that the rat made a response to R (L) in trial $i-j$ regardless of the stimulus and 0 means the rat made an abort. The response after correct (incorrect) choices predictor was identical to the response predictor, except that the indicator was set to zero when the trial was either an abort or an incorrect (correct) trial.

For each of these four aspects, we fit a kernel that extended 15 trials into the past, which was convolved with the corresponding predictor. We represented all history kernels as linear combinations of five unit-norm decaying exponentials with time constants of [0.5, 1, 2, 4, 8] trials (Supplementary Fig. 10g, inset). In practice, we used an orthonormal basis in the space spanned by these five exponentials. Thus, all in all, our model fits 32

coefficients plus the bias: 12 from predictors associated to the current trial and 5 coefficients for each of the 4 history kernels. All coefficients and kernels for all animals are shown in Supplementary Fig. 10.

Model fits were performed using the Matlab version of the free software package glmnet[53] (https://web.stanford.edu/~hastie/glmnet_matlab/). Regularization was implemented using the elastic net method with parameter $\alpha = 0.5$. All predictors were subject to regularization. All fits used cross-validation as implemented by the function cvglmnet and we used the lamda_min option to select the hyper-parameter that minimizes prediction error. In addition, for assessing the predictive power of the fits, we manually implemented nested cross-validation with five outer folds. Briefly, model coefficients and hyper-parameters were sequentially fit using four-fifths of the data and prediction was evaluated on the remaining one-fifth until all the data had been used both for fitting and prediction. Predictive quality was assessed using the receiver operating characteristic area under the curve (AUC; Fig. 5c, inset).

To estimate the relative predictive power of different types of predictors, we used the relative variance of the linear component of the model for a particular set of predictors. After linear combination or convolution for both current-trial coefficients and the four different history kernels (best fits), we obtained five time series across trials. For obtaining a final predictive probability, these would be added. Instead, we computed the variance across trials associated with each of the five time series and report its value relative to the summed variance across the five. The fraction of variance for each of the five type of predictors for each rat are shown in Fig. 5c and Supplementary Fig. 10.

We used $n = 2{,}000$ bootstrap re-samples to estimate uncertainty (95% CI, shown as error bars in Fig. 5c and Supplementary Fig. 10) on the model parameters[54].

To assess whether trial history made a significant contribution to the predictive power of the model, we compared the AUC of the actual data to that of shuffled surrogate datasets (Fig. 5c, inset). In each shuffled surrogate dataset, history predictors were randomly permuted across the whole set of trials, with the constraint that the four

history predictors associated with a particular trial were in the same position of the permuted sequence for the surrogate dataset. For each surrogate, we estimated the uncertainty in the coefficients again using bootstrapping. Trial history was deemed to make a significant contribution if there was no overlap between the 95% CI of the real data and the shuffled surrogates.

Model fitting

Our model is specified by the following four parameters: θ_e, λ, T_0 (the inverse of the effective gain of the pressure-to-rate transformation Nr_0) and t_{NDT} (no parameters describing lapses or biases are necessary to model the data).

Constrained model fitting approach

To test the whether the speed–accuracy trade-off as a function of ILD and whether the dependency of RTs on ABL, which are structural features of the model, were also present in the data, we used an unorthodox 'constrained' model fitting approach. Here, we first used choice data to fit the single parameter Γ (equation (2)) from the logistic function as follows:

$$\text{Choose} - \text{rightprobability} \equiv p_+ = \frac{1}{1 + \exp(-2\Gamma \text{ILD})}$$

Next, we fit the RTDs. We only fit the RTDs of correct trials. Although we also attempted to fit incorrect trials, the RTDs for incorrect trials, especially for low ABLs, were severely distorted due to the anticipation by the rats of the sound onset discussed in the main text and in Supplementary Fig. 2. We could not find a principled way of finding and excluding anticipatory error RTs that would lead to well behaved RTDs for errors. In the constrained fitting approach, we considered Γ was fixed (to its estimated value from fits to accuracy) for fits of the RTDs (implying that all dependence of the RT distributions on ILD was not available for the model fitting process to modify). To further test the ABL-to-RT relationship predicted by the model, we excluded data from the $\text{ABL} = 40$ dB SPL condition for these fits. RTDs were fit using the χ^2 method

because of its robustness and efficiency[55]. In this method, one uses the χ^2 statistic for a dataset composed of measurements of N_p proportions. We used ten proportions (split by the 0.1, 0.2 and 0.9 quantiles of the RT distribution) per condition (although in Fig. 4c and Supplementary Fig. 7b1–5, we only show the 0.1, 0.3 and 0.9 quantiles to avoid visual clutter), and considered eight conditions (the four difficulties for ABL = 20 and 60 dB SPLs). Thus, in the constrained approach, we adjust the value of three free parameters (λ, T_0 and t_{NDT}) to fit the results for $N_p = 80$ proportions/data points. As discussed above, only RTs > RT_{min} and RTs < RT_{max} were considered. The cost function for the fit was as follows:

$$\chi^2 = \sum_{1}^{N_p} \frac{(O - E)^2}{E}$$

where O is the observed proportions and E represents the expected proportions given the model. To estimate E, we used the two series approximations of the RTD of a drift-diffusion process $\rho_{DT}(t)$ for short and long decision-times[56,57], truncated both series at five terms, and selected between the two according to a threshold in dimensionless time of 0.25, which guarantees a smooth overlap and ensures that the relative error of the approximation is always below 10^{-3} for all decision times. The distributions were fit in dimensionless time. For a given non-decision time t_{NDT} and measured RT, the dimensionless decision-time is $\tau_{DT} = (RT - t_{NDT})/t_\theta$, where t_θ is defined in equation (3). The χ^2 cost function was minimized numerically in Matlab with the function fminsearch, which uses an implementation of the Nelder–Mead simplex algorithm[58]. The initial conditions for each parameter were randomly selected within a reasonable interval each time the function was executed, always converging to the same global minimum. Since our empirical RTDs are only defined between RT_{min} and RT_{max} (see above), for each putative value of the model parameters, the model RTD was compared with empirical RTD only within this range to derive the corresponding value of the cost. The purpose of this procedure is to avoid selecting parameter values that attempt to accommodate the discontinuities in the empirical RTDs resulting from the truncation process by changing the shape of non-truncated model distributions.

RTD fits were performed for both individual rats and pooled data across rats. When fitting pooled data, we first calculated the quantiles of the RTDs for each rat separately, then we averaged them (weighted by the relative number of trials performed by each rat), and then we fit the model to the average quantiles.

Unconstrained model fitting approach

We also fit the data in a standard unconstrained fashion. In this approach, we also used the χ^2 method, but to simultaneously fit accuracy and RT, we scaled the RTDs so that the area under each distribution would be equal to the probability of the corresponding choice given the model. Accordingly, for each condition, we used 11 proportions, ten for the RTs of correct trials, and a single one for the incorrect trials (in this way, incorrect trials are only used for the estimation of accuracy, see above). Thus, in this approach, we adjust the value of all four free parameters to fit the results for $N_p = 132$ proportions (11 per condition for 12 conditions). All other aspects of the χ^2 fit were the same as for the constrained approach.

To obtain error bars for our estimates and an estimate of the joint distribution of the parameters from the fit (Fig. 4c; Supplementary Figs. 7–9; Supplementary Table 3), we used bootstrapping[52], performing subsequent fits for $N_r = 1{,}000$ re-samples (with replacement) from our dataset.

Maximum likelihood fits

To ensure that our results were robust against the model fitting method, both constrained and unconstrained approaches were used using the quantile maximum likelihood method[59]. This method maximizes the likelihood of the data under the model, but also grouping RTs into quantiles to minimize problems due to outliers. The results of the fits using this method (reported in Supplementary Table 3) are effectively identical to those obtained using the χ^2 method.

Multiplicative noise of baseline firing rates

We also fitted the data with an extended model that considers multiplicative variability of the baseline firing rate of each channel of sensory neurons (Supplementary Note). The effect of the additional parameter σ_r was implemented in our custom Matlab fitting algorithm through numerical integration with the function integral2. We used the unconstrained χ^2 method and proceeded as otherwise described above. Parameter uncertainties were again estimated using bootstrapping (700 re-samples).

Data availability

All data are available from the authors upon reasonable request.

Code availability

The custom Matlab scripts used to analyze the data are available upon reasonable request.

References

1. Weber, E. H. *De Pulsu, Resorptione, Auditu et Tactu: Annotationes Anatomicae et Physiologicae, Auctore* (Prostat Apud CF Koehler, 1834).
2. Laming, D. *Sensory Analysis* (Academic, 1986).
3. Fechner, G. T. *Element der Psychophysik* (Breitkopf and Harterl, 1860).
4. Link, S. W. *The Wave Theory of Difference and Similarity* (Psychology Press, 1992).
5. Gescheider, G. A. *Psychophysics: the Fundamentals* (Psychology Press, 2013).
6. Treisman, M. Noise and Weber's law: the discrimination of brightness and other dimensions. *Psychol. Rev.* **71**, 314–330 (1964).
7. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, 1966).
8. Deco, G. & Rolls, E. T. Decision-making and Weber's law: a neurophysiological model. *Eur. J. Neurosci.* **24**, 901–916 (2006).
9. Teodorescu, A. R., Moran, R. & Usher, M. Absolutely relative or relatively absolute: violations of value invariance in human decision making. *Psychon. Bull. Rev.* **23**, 22–38 (2016).
10. Simen, P., Vlasov, K. & Papadakis, S. Scale (in)variance in a unified diffusion model of decision making and timing. *Psychol. Rev.* **123**, 151–181 (2016).

11. Henmon, V. A. C. The relation of the time of a judgment to its accuracy. *Psychol. Rev.* **18**, 186–201 (1911).
12. Wald, A. & Wolfowitz, J. Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* **19**, 326–339 (1948).
13. Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
14. Forstmann, B. U., Ratcliff, R. & Wagenmakers, E.-J. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annu. Rev. Psychol.* **67**, 641–666 (2016).
15. Wesolek, C. M., Koay, G., Heffner, R. S. & Heffner, H. E. Laboratory rats (*Rattus norvegicus*) do not use binaural phase differences to localize sound. *Hear. Res.* **265**, 54–62 (2010).
16. Grothe, B., Pecka, M. & McAlpine, D. Mechanisms of sound localization in mammals. *Physiol. Rev.* **90**, 983–1012 (2010).
17. Stellmack, M. A., Viemeister, N. F. & Byrne, A. J. Monaural and interaural intensity discrimination: level effects and the binaural advantage. *J. Acoust. Soc. Am.* **116**, 1149–1159 (2004).
18. Recanzone, G. H. & Beckerman, N. S. Effects of intensity and location on sound location discrimination in macaque monkeys. *Hear. Res.* **198**, 116–124 (2004).
19. Nodal, F. R., Bajo, V., Parsons, C. H., Schnupp, J. W. & King, A. J. Sound localization behavior in ferrets: comparison of acoustic orientation and approach-to-target responses. *Neuroscience* **154**, 397–408 (2008).
20. Mendonça, A. et al. The impact of learning on perceptual decisions and its implication for speed–accuracy tradeoffs. Preprint at *biorXiv* https://doi.org/10.1101/501858 (2018).
21. Gillespie, D. T. *Markov Processes: an Introduction for Physical Scientists* (Elsevier, 1991).
22. Ratcliff, R. & Rouder, J. N. Modeling response times for two-choice decisions. *Psychol. Sci.* **9**, 347–356 (1998).
23. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108**, 550–592 (2001).
24. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
25. Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N. & Pouget, A. The cost of accumulating evidence in perceptual decision making. *J. Neurosci.* **32**, 3612–3628 (2012).
26. Kiani, R., Hanks, T. D. & Shadlen, M. N. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* **28**, 3017–3029 (2008).
27. Brunton, B. W., Botvinick, M. M. & Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340**, 95–98 (2013).
28. Hartmann, W. M. & Constan, Z. A. Interaural level differences and the level-meter model. *J. Acoust. Soc. Am.* **112**, 1037–1045 (2002).
29. Greene, N. T. et al. Spatial hearing ability of the pigmented guinea pig (*Cavia porcellus*): minimum audible angle and spatial release from masking in azimuth. *Hear. Res.* **365**, 62–76 (2018).

30. Matthews, W. J., Stewart, N. & Wearden, J. H. Stimulus intensity and the perception of duration. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 303–313 (2011).
31. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).
32. Tollin, D. J., Koka, K. & Tsai, J. J. Interaural level difference discrimination thresholds for single neurons in the lateral superior olive. *J. Neurosci.* **28**, 4848–4860 (2008).
33. Jones, H. G., Brown, A. D., Koka, K., Thornton, J. L. & Tollin, D. J. Sound frequency-invariant neural coding of a frequency-dependent cue to sound source location. *J. Neurophysiol.* **114**, 531–539 (2015).
34. Tsai, J. J., Koka, K. & Tollin, D. J. Varying overall sound intensity to the two ears impacts interaural level difference discrimination thresholds by single neurons in the lateral superior olive. *J. Neurophysiol.* **103**, 875–886 (2009).
35. Kyweriga, M., Stewart, W. & Wehr, M. Neuronal interaural level difference response shifts are level-dependent in the rat auditory cortex. *J. Neurophysiol.* **111**, 930–938 (2013).
36. Benichoux, V., Brown, A. D., Anbuhl, K. L. & Tollin, D. J. Representation of multidimensional stimuli: quantifying the most informative stimulus dimension from neural responses. *J. Neurosci.* **37**, 7332–7346 (2017).
37. Stecker, G. C., Harrington, I. A. & Middlebrooks, J. C. Location coding by opponent neural populations in the auditory cortex. *PLoS Biol.* **3**, e78 (2005).
38. Keating, P., Dahmen, J. C. & King, A. J. Complementary adaptive processes contribute to the developmental plasticity of spatial hearing. *Nat. Neurosci.* **18**, 185–187 (2015).
39. Gibbon, J. Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* **84**, 279–325 (1977).
40. Simen, P. et al. A model of interval timing by neural integration. *J. Neurosci.* **31**, 9238–9253 (2011).
41. Scott, B. B., Constantinople, C. M., Erlich, J. C., Tank, D. W. & Brody, C. D. Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats. *eLife* **4**, e11308 (2015).
42. Tsetsos, K., Usher, M. & McClelland, J. L. Testing multi-alternative decision models with non-stationary evidence. *Front. Neurosci.* **5**, 63 (2011).
43. Wagenmakers, E.-J. & Brown, S. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychol. Rev.* **114**, 830–841 (2007).
44. Zwislocki, J. J. in *Sensation and Measurement* (eds. Moskowitz, H. R., Scharf, B. & Stevens, J. C.) 185–197 (Springer, 1974).
45. Martin, P. & Hudspeth, A. Compressive nonlinearity in the hair bundle's active response to mechanical stimulation. *Proc. Natl Acad. Sci. USA* **98**, 14386–14391 (2001).
46. Stevens, S. S. On the psychophysical law. *Psychol. Rev.* **64**, 153–181 (1957).
47. Dehaene, S. The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* **7**, 145–147 (2003).

48. Uchida, N. & Mainen, Z. F. Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.* **6**, 1224–1229 (2003).
49. Brown, A. D., Benichoux, V., Jones, H. G., Anbuhl, K. L. & Tollin, D. J. Spatial variation in signal and sensory precision both constrain auditory acuity at high frequencies. *Hear. Res.* **370**, 65–73 (2018).
50. Uchida, N., Kepecs, A. & Mainen, Z. F. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nat. Rev. Neurosci.* **7**, 485–491 (2006).
51. Loader, C. *Local Regression and Likelihood* (Springer Science & Business Media, 2006).
52. Friedman, J, Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2001).
53. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
54. Hastie, T., Tibshirani, R. & Wainwright, M. *Statistical Learning with Sparsity: the Lasso and Generalizations* (CRC Press, 2015).
55. Ratcliff, R. & Tuerlinckx, F. Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.* **9**, 438–481 (2002).
56. Blurton, S. P., Kesselmeier, M. & Gondan, M. Fast and accurate calculations for cumulative first-passage time distributions in wiener diffusion models. *J. Math. Psychol.* **56**, 470–475 (2012).
57. Gondan, M., Blurton, S. P. & Kesselmeier, M. Even faster and even more accurate first-passage time densities and distributions for the wiener diffusion model. *J. Math. Psychol.* **60**, 20–22 (2014).
58. Lagarias, J. C., Reeds, J. A., Wright, M. H. & Wright, P. E. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* **9**, 112–147 (1998).
59. Heathcote, A. & Brown, S. Reply to Speckman and Rouder: a theoretical basis for QML. *Psychon. Bull. Rev.* **11**, 577–578 (2004).

Acknowledgements

The authors thank D. Reato, J. de la Rocha, G. Agarwal, E. Lottem, A. Nevado, J. Paton and L. Petreanu for careful reading of the manuscript. They also thank P. Simen for pointing out relevant references, T. Stensola for help with behavioral boxes, the vivarium platform at Champalimaud Research for support, M. Bayonas for help with headphone prototyping and D. Kobak for reading the manuscript and advice on statistical analyses. J.L.P.-V. was supported by the HFSP postdoctoral scholarship LT 000442/2012, and J.R.C.-S., M.V., T.C., M.I.V. and A.G.M. were supported by doctoral fellowships from the Fundação para a Ciência e a Tecnologia. Z.F.M. was supported by the Champalimaud Foundation, the European Research Council (Advanced Investigator

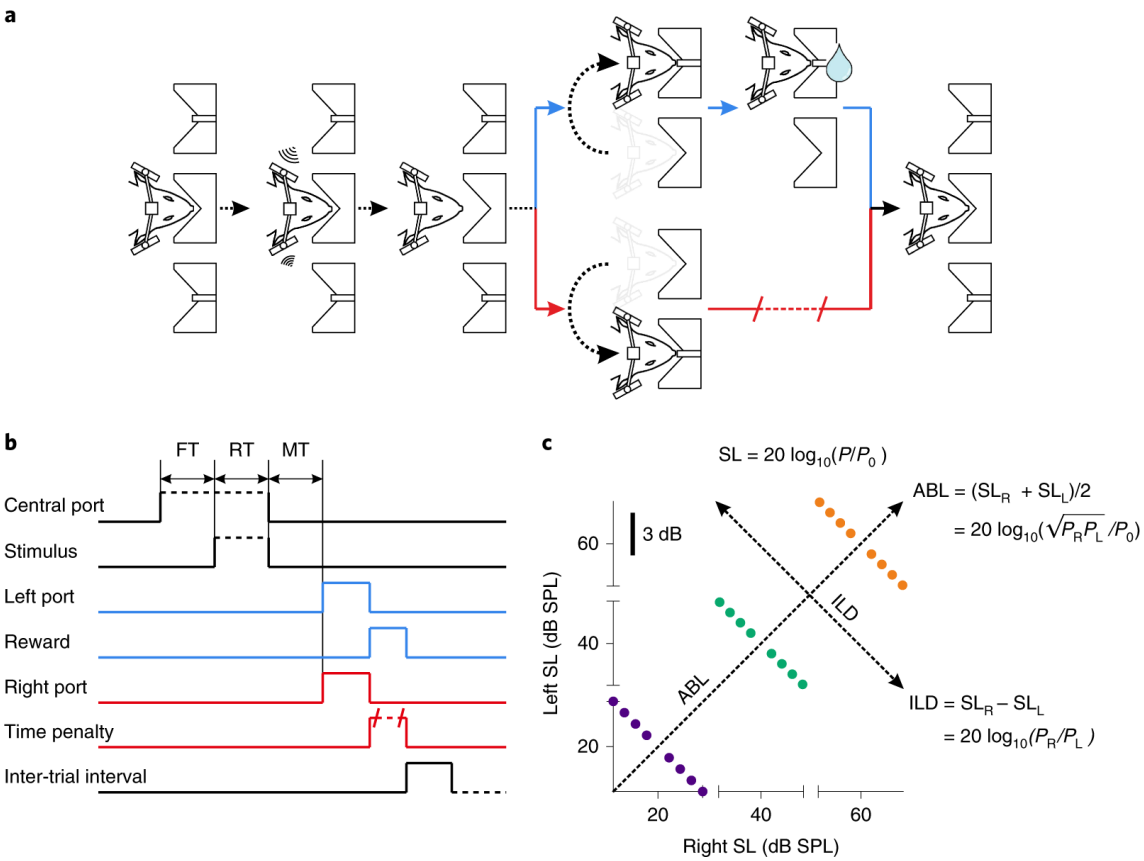
grant 250334), the Human Frontier Science Program (grant RGP0027/2010) and the Simons Foundation (grant 325057). A.R. was supported by the Champalimaud Foundation, a Marie Curie Career Integration grant (PCIG11-GA-2012-322339), the HFSP Young Investigator Award RGY0089 and the EU FP7 grant ICT-2011-9-600925 (NeuroSeeker).

Ethics declarations

Competing interests

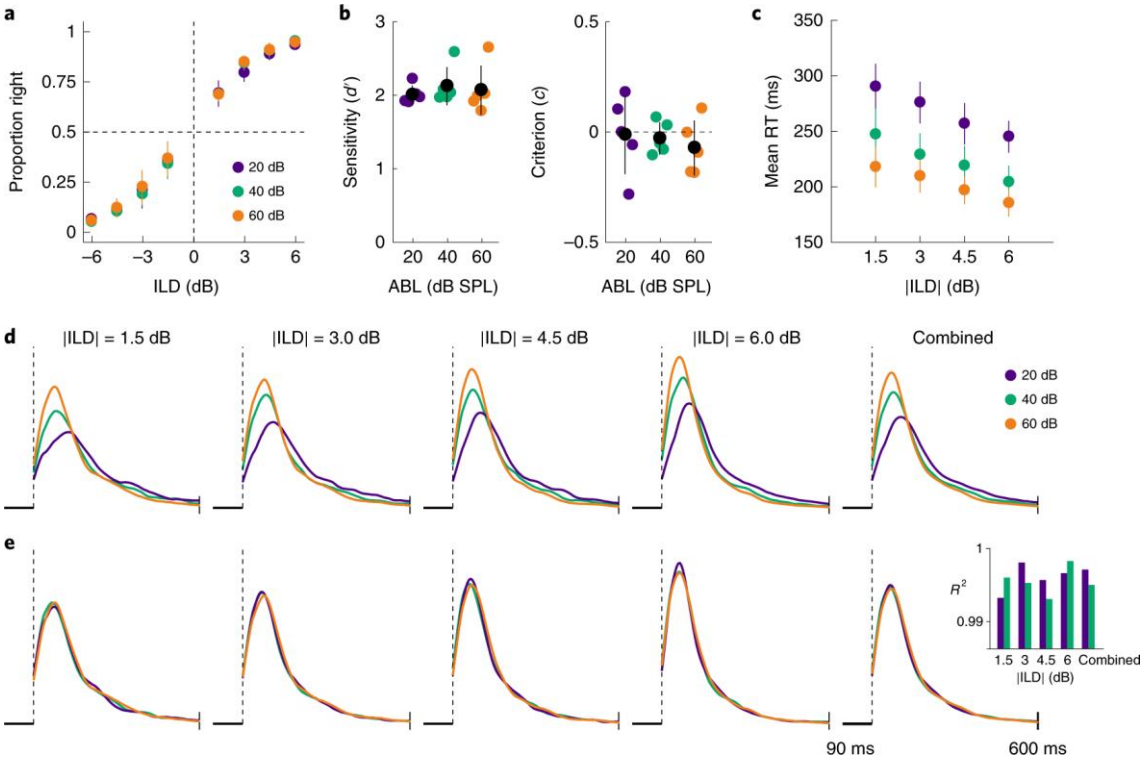
The authors declare no competing interests.

Fig. 1: Task structure and stimulus set



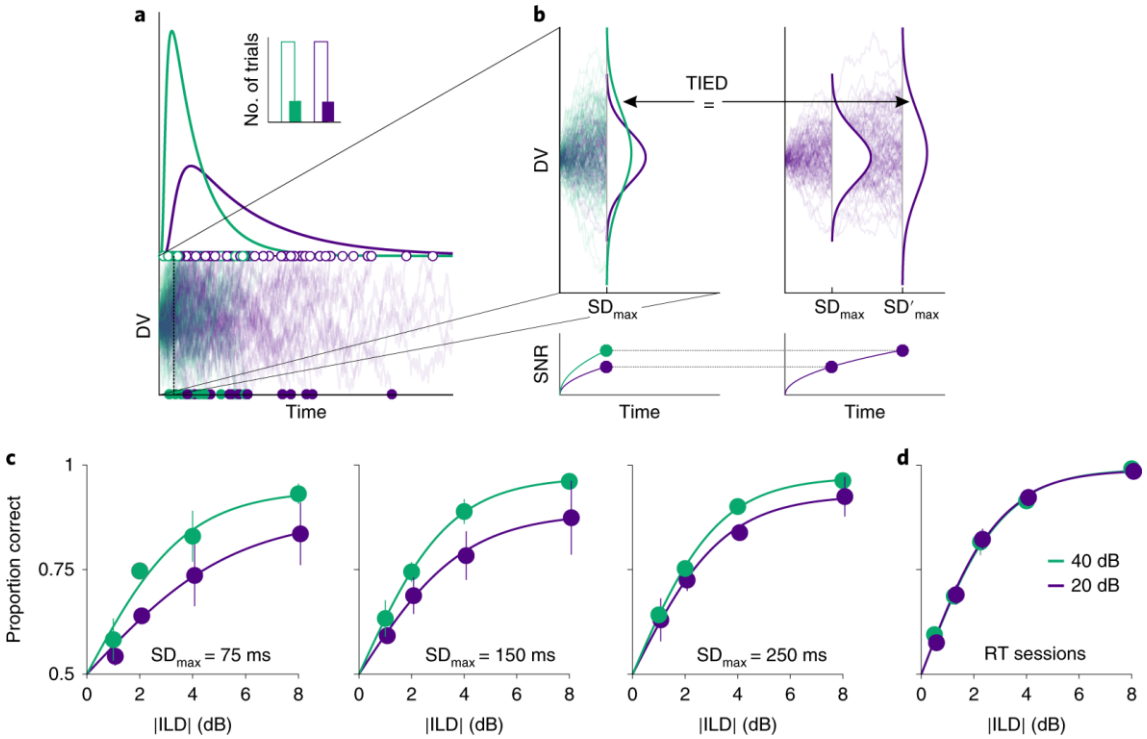
a, Schematic depiction of the different task events. Rats were rewarded with water for making the correct choice (blue) and were punished with a time delay for making an error (red). **b**, Timeline of the relevant task events. FT, fixation time; MT, movement time. **c**, Stimulus set. The ABL (ILD) of a particular stimulus is given by the average (difference, by convention right minus left) of the intensity of the sound in dB SPL (sound level (SL)) across both speakers. Constant ILD (ABL) implies constant intensity ratios (intensity products). $P_0 = 20$ μPa is the reference pressure of the dB SPL scale. Purple, green and orange circles indicate ABLs of 20, 40 and 60 dB SPL, respectively.

Fig. 2: Time–intensity equivalence in sensory discrimination



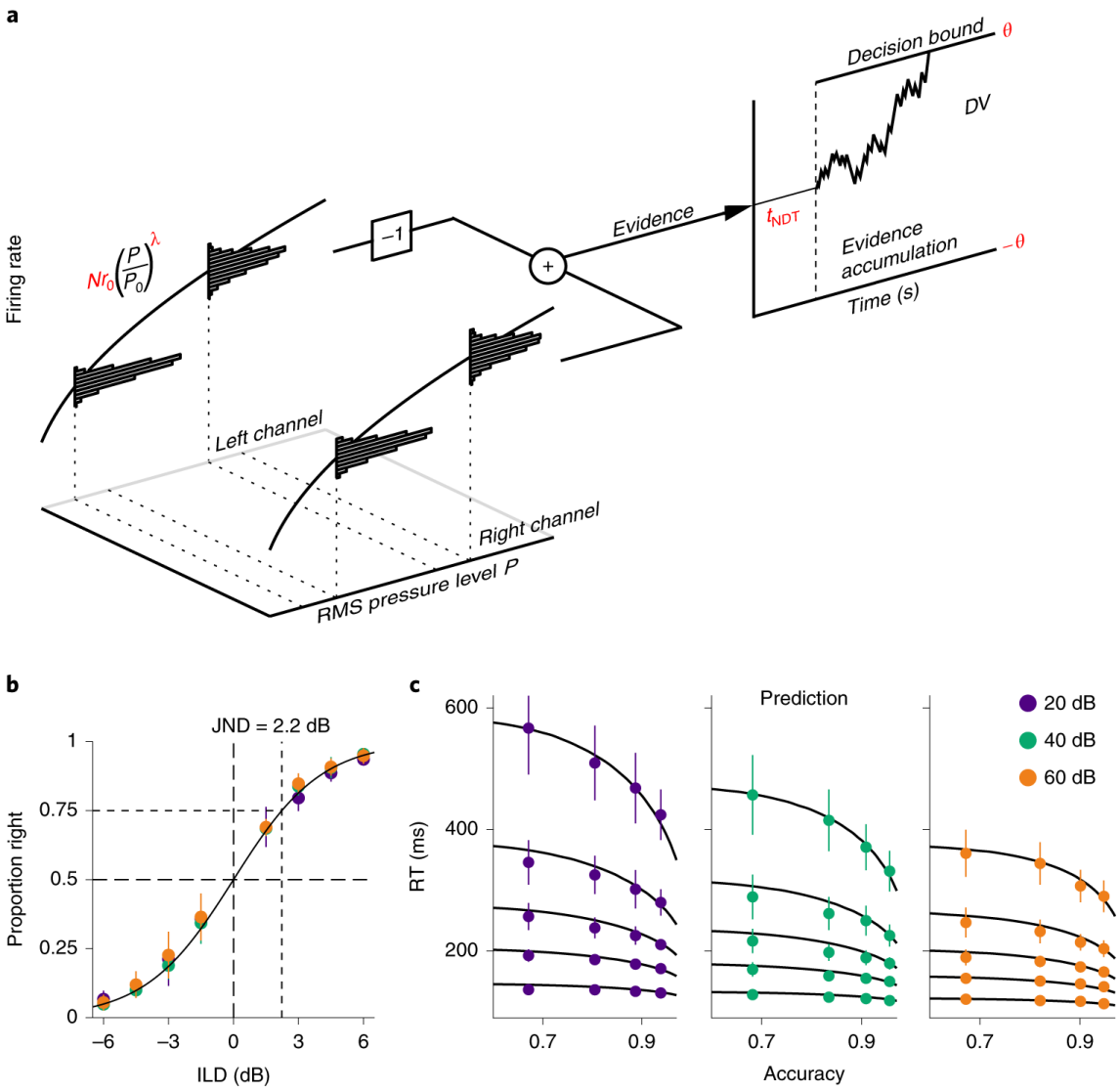
a, Choose-right probabilities as a function of ILD for each ABL separately (mean ± s.d. across rats, $n = 5$ rats). **b**, Sensitivity (d') and criterion (c) for each animal (black filled circles represent the mean; error bars are ±1 s.d. across rats). **c**, RT (mean ± s.e.m. across rats) as a function of difficulty for each ABL separately. **d**, RTDs for the three ABLs are shown separately for each difficulty, and combined across difficulties (far right). For all RTDs, the broken line indicates the time at which RTs become condition dependent (scale bar in all plots, 90 ms; see Methods and Supplementary Fig. 2). Each RTD contains all data for that condition from all rats. **e**, For each difficulty, we uniformly rescaled the time to maximize the overlap of each RTD with that for ABL = 60 dB SPL (Methods and Supplementary Fig. 3). The inset shows the fraction of variance (R^2) that the RTD at ABL = 60 dB explains about the rescaled RTDs at 20 and 40 dB SPLs (Methods and Supplementary Fig. 3).

Fig. 3: Breakdown of WL for controlled short sound durations



a, Schematic illustration of the TIED. Sample paths of the DV for two stimuli with a high (green) and low (purple) ABL but the same ILD are shown. Horizontal lines are the decision bounds. Open circles mark the RT for correct trials, while filled circles mark incorrect trials. The RTDs for the two stimuli are plotted on the top. Because the green stimulus is stronger, the green sample paths rise faster, but if the four conditions we identify are met, the proportion of trials hitting either bound is ABL-independent (inset) and the two RTDs are related to each other by a rigid stretching of time. **b**, Schematic illustration of prediction. For times that are short compared with the typical RTs (rectangle outlined with a broken line in **a**), the likelihood of hitting either bound is still low and one can focus on the dynamics of the DV ignoring the bound. Left: sample paths and distribution of the DV at the putative offset SD_{max} of the same two sounds in **a**. Lower left: temporal evolution of the signal-to-noise ratio (SNR) of the DV for the two stimuli up to that time. Right: the TIED implies that the distribution of the DV for the green (loud) stimulus at SD_{max} should be identical to the distribution of the DV for the purple (quiet) stimulus at a later time SD′maxSDmax′. Lower right: since the SNR of the subthreshold DV always increases with time under evidence accumulation, performance for the weaker (purple) stimulus at SD_{max} is expected to be worse. **c**, Percent correct as a function of difficulty (mean ± s.d. across rats, $n = 4$ rats) for ABL = 20 and 40 dB SPLs. **d**, Same as for **c**, but for the RT sessions.

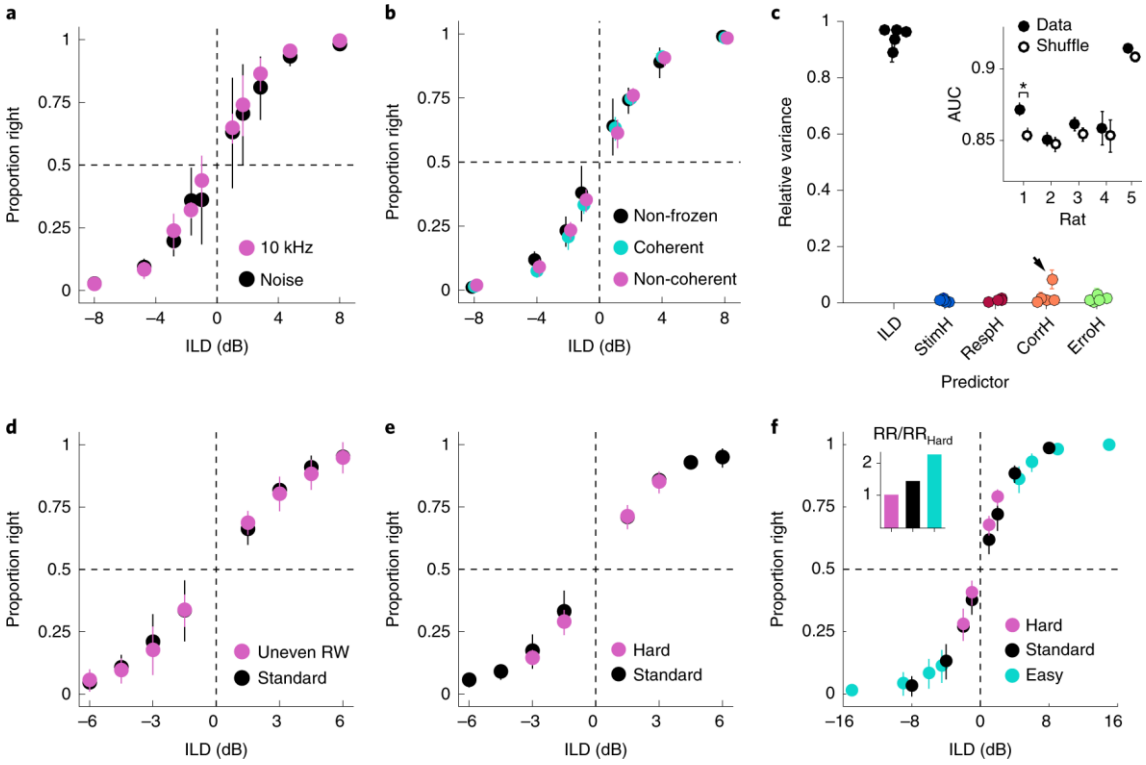
Fig. 4: Modeling the effect of difficulty and overall intensity in sensory discrimination



a, Model structure and parameters. Each stimulus is encoded through a compressive power-law non-linearity of gain r_0 and exponent λ by a population of N neurons (a sensory channel) firing with Poisson statistics. The sensory evidence is the difference between the activity of each channel. The DV integrates the evidence until a constant bound at $\pm\theta$ is hit (decision time). RTs are the sum of the decision time and a stimulus-independent non-decision time t_{NDT}. The four model parameters are shown in red. **b**, Choose-right probabilities (circles, mean ± s.d. across rats, $n = 5$ rats) and fit to these data from the single-parameter model in equation (2) (line). **c**, Circles show the 0.1, 0.3, 0.5, 0.7 and 0.9 quantiles (mean ± s.e.m. across rats) of the measured RTDs. Black lines are the results from the model. For all panels, the curvature and relative position of the model quantiles (black lines) are a prediction derived from the fit to accuracy. For the left and right panels, a single vertical offset (t_{NDT}) and a rigid vertical

stretching of the quantiles (t_θ(ABL)) are chosen to match the data. The middle panel is a full prediction. See Supplementary Fig. 7 for similar fits for each individual rat.

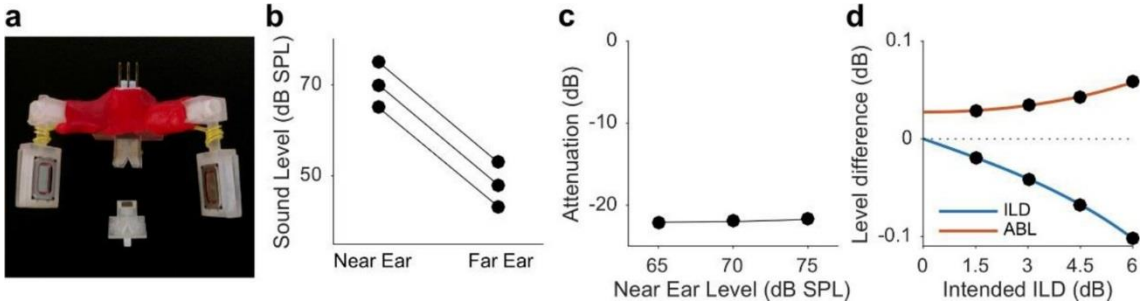
Fig. 5: Behavioral manipulations



a, Choose-right probability (mean ± s.d. across rats, $n = 3$ rats) for broadband noise and 10 kHz pure tones of ABL = 60 dB SPL. **b**, Same for the standard non-frozen condition and for frozen noise either coherent or incoherent across ears (mean ± s.d. across rats, $n = 6$; Methods). **c**, History effects using logistic regression. The fraction of variance of the linear part of the model explained by the stimulus (ILD), stimulus history (StimH), response history (RespH), correct trials (CorrH) and response after errors (ErroH) (Methods and Supplementary Fig. 10). Each circle represents a single rat ($n = 5$ rats). The arrow for CorrH is rat 1. The inset shows the prediction accuracy (using nested cross-validation; Methods) of the logistic regression for the actual data and for surrogate data for which the history information was shuffled. For this panel and its inset, error bars are bootstrap 95% CI (2,000 re-samples) and the asterisk indicates that the actual and surrogate CIs do not overlap. **d**, Choose-right probabilities (mean ± s.d. across rats, $n = 4$ rats) for standard blocks and for uneven reward (RW) blocks, in which rewards for correct choices in the two hardest (easiest) conditions are 20% larger (smaller). **e**, Same ($n = 3$ rats) for hard versus control blocks. **f**, Same for more extreme, bidirectional and longer-lasting manipulations of motivation ($n = 5$ rats; see Methods for details on all task manipulations; Supplementary Table 2) The inset shows the reward rate (RR) in each condition of motivation relative to the reward rate in the hard condition (RR_{Hard}).

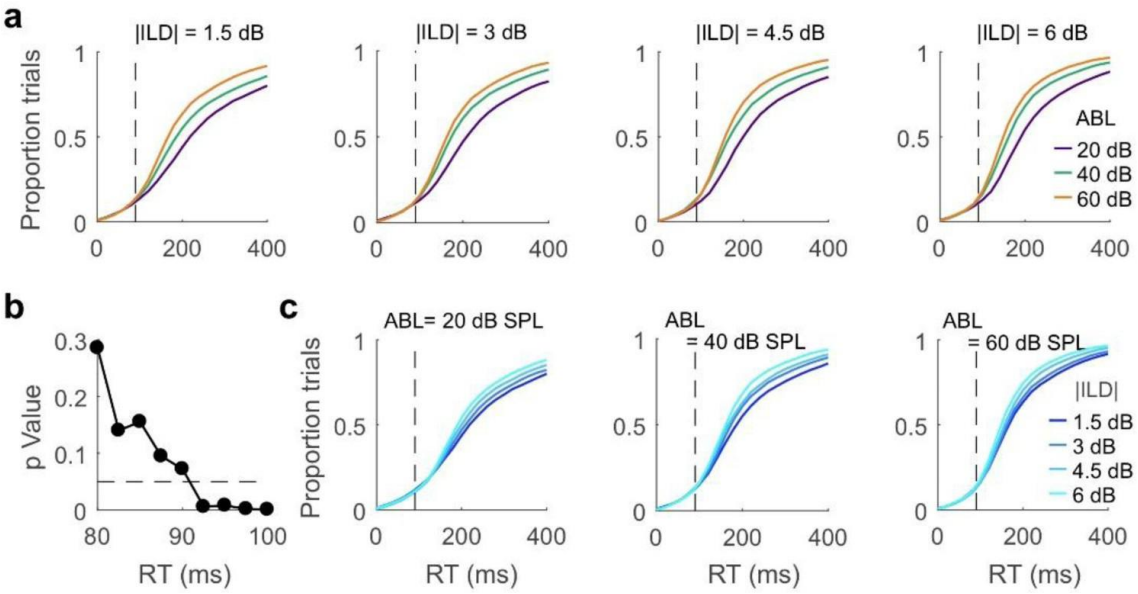
Integrated supplementary information

Supplementary Figure 1 Headphones and acoustic shadow of the head**.**



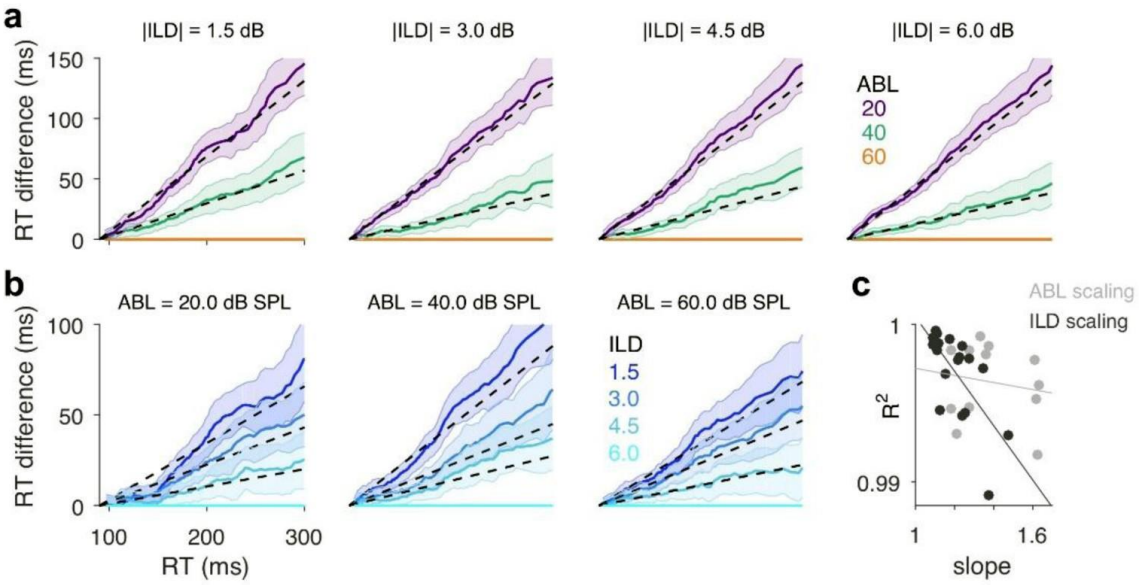
(**a**) Headphones used for stimulus delivery, and base (bottom). The base is chronically implanted to the skull of the animals and contains a strong miniature magnet which both holds the speakers in place during task performance and also allows easy attachment/detachment. The actual positioning of the speakers was adjusted under anesthesia to match the position of the pinnae for each rat individually relative to the base. The red material in the picture is a moldable glue (Sugru) used for strengthening the headphone structure. **(b)** Measurements of the acoustic shadow generated by the head. Broadband noise stimuli were played at 65, 70 and 75 dB SPL and sound level was measured placing the microphone by the ear canal of the 'far' ear. **(c)** The head plus near field positioning of the speakers causes an attenuation of 22 dB. (**d**) Because the head attenuation is not infinite, the sound at each ear contains a contribution from the intensity of both speakers. Thus, the intended (using the sound intensity only from the near speaker) and actual experienced intensity at each ear are not identical. We calculated the difference between the actual and experienced intensity assuming additivity of the squared pressure RMS from each speaker, and we used this to compute the difference between actual and intended ILD and ABL, which are shown in panel (d) as a function of intended ILD. Since these differences are always less than 0.1 dB (which is less than an order of magnitude smaller than the just noticeable difference (JND) for ILD of our animals) we have, for simplicity, ignored the difference between actual and intended levels throughout this study.

Supplementary Figure 2 Short RTs and processing delays**.**



(a) Each plot shows the cumulative distribution of RTs for a given difficulty (absolute value of ILD) for the three ABLs (RTs merged across rats). The shape of the distribution for the shorter RTs is the same across all conditions. This is presumably due to the fact that sound onset after central-port entry was to some extent predictable (fixation period was drawn from a uniform distribution between 300 and 350 ms) and to the fact that we did not impose a minimum RT. These short, condition-independent RTs are thus the result of anticipation. (**b**) To detect the earliest time where RTs become condition-dependent, we run a two tailed Kolmogorov-Smirnoff test comparing the RTs for the fastest and slowest conditions (|ILD| = 6 dB and ABL = 60 dB SPL versus (|ILD| = 1.5 dB and ABL = 20 dB SPL) as a function of the maximum RT included in the comparison. This panel shows the p-value of the test. We defined RT_{min} as the value at which this comparison becomes significant (90 ms, $p<0.05$). This value is shown as a dashed vertical line in (a) and (c). Since we are interested in the stimulus-dependence of RTs, we excluded from all analyses RTs smaller than RT_{min}. **(c)** Same as (a) but comparing the cumulative RTDs across ILDs for each ABL separately. As expected, the distributions start to diverge later for fainter sounds. However, it is obvious that this unspecific intensity-dependent delay cannot account for the changes in the RTD as a function of ABL shown in (a). None of our results changes qualitatively if we define a separate RT_{min} for each ABL.

Supplementary Figure 3 Empirical assessment of RT scale invariance.

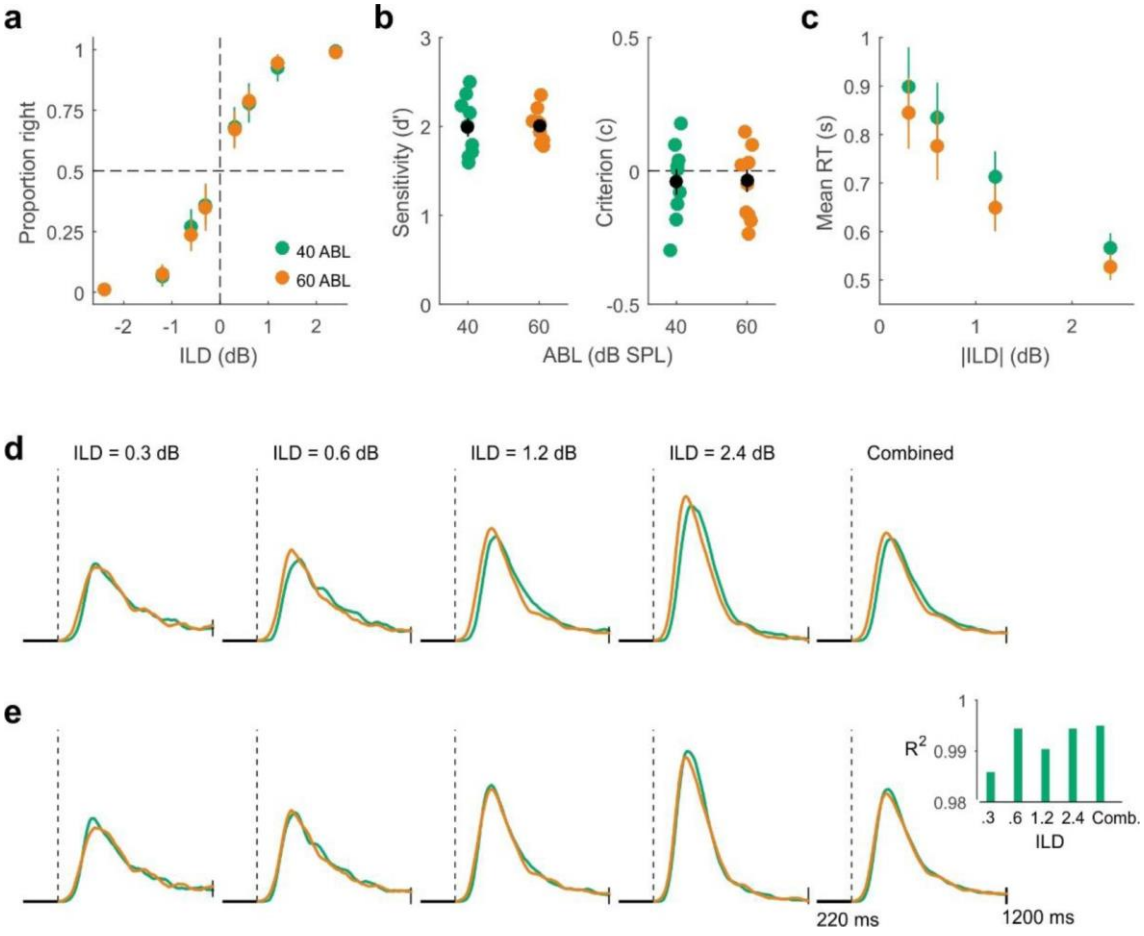


When two distributions in time are related to each other through a rigid stretching of the time axis, their percentiles are proportional. To see this, assume that distribution ρ'(t) is obtained from ρ(t) by stretching of the time axis by a factor α, so that ρ(t) = α ρ' (αt). It follows that.

$\int_0^{\tau} dt\rho(t) = \alpha \int_0^{\tau} dt\rho'(\alpha t) = \int_0^{\alpha\tau} dt'\rho'(t')$ for all τ⊡ If the value of the previous integrals is Q/100, then τ is the Q^{th} percentile $PC_{\rho}(Q)$ of ρ(t) and ατ is the Q^{th} percentile $PC_{\rho'}(Q)$ of ρ'(t), i.e., their percentiles are proportional. This fact allows us to use the slope of a linear fit of the percentiles of the two distributions to identify the (putative) temporal rescaling factor α corresponding to a given change in ABL, and the R^2 of the fit to quantify its precision. **(a)** Each plot shows, for each difficulty, $PC_{ABL}(Q)$-$PC_{60dB}(Q)$ as a function of $PC_{60\ dB}(Q)$, where $PC_{ABL}(Q)$ is the $100Q^{th}$ percentile of the RT distribution for the corresponding ABL of that difficulty. Each RTD contains all data for that condition from all rats (n = 5 rats). Although, for each pair of RTDs, the linear fit was made directly between their percentiles, we plot in the figure the difference between the two percentiles against one of them, because it graphically displays the linear relationship more clearly. Only RTs > RT_{min} are considered in this analysis. Dashed line is the outcome of the linear fit. Shaded regions represent three standard deviations of the sampling distribution of the percentiles (see Methods). Note that higher percentiles have more error. Because of this, we used weighted least squares to do the linear fit. **(b)** Since difficulty has been suggested to also lead to an approximate rescaling of the RTD (Wagenmakers, E. and Brown, S., *Psychological Rev.*, 114(3):830, 2007; Ratcliff, R. and McKoon, G., *Neural Computation*, 20(4):873-922, 2008; Srivastava, V., et al., *Journal of Mathematical Psychology*, 75:96-109, 2016), we performed the same analysis as in (a) (n = 5 rats) for pairs of distributions corresponding to different difficulties (for each ABL separately). **(c)** R^2 of the weighted least squares fit vs slope of the fit for the effect of changes on ABL

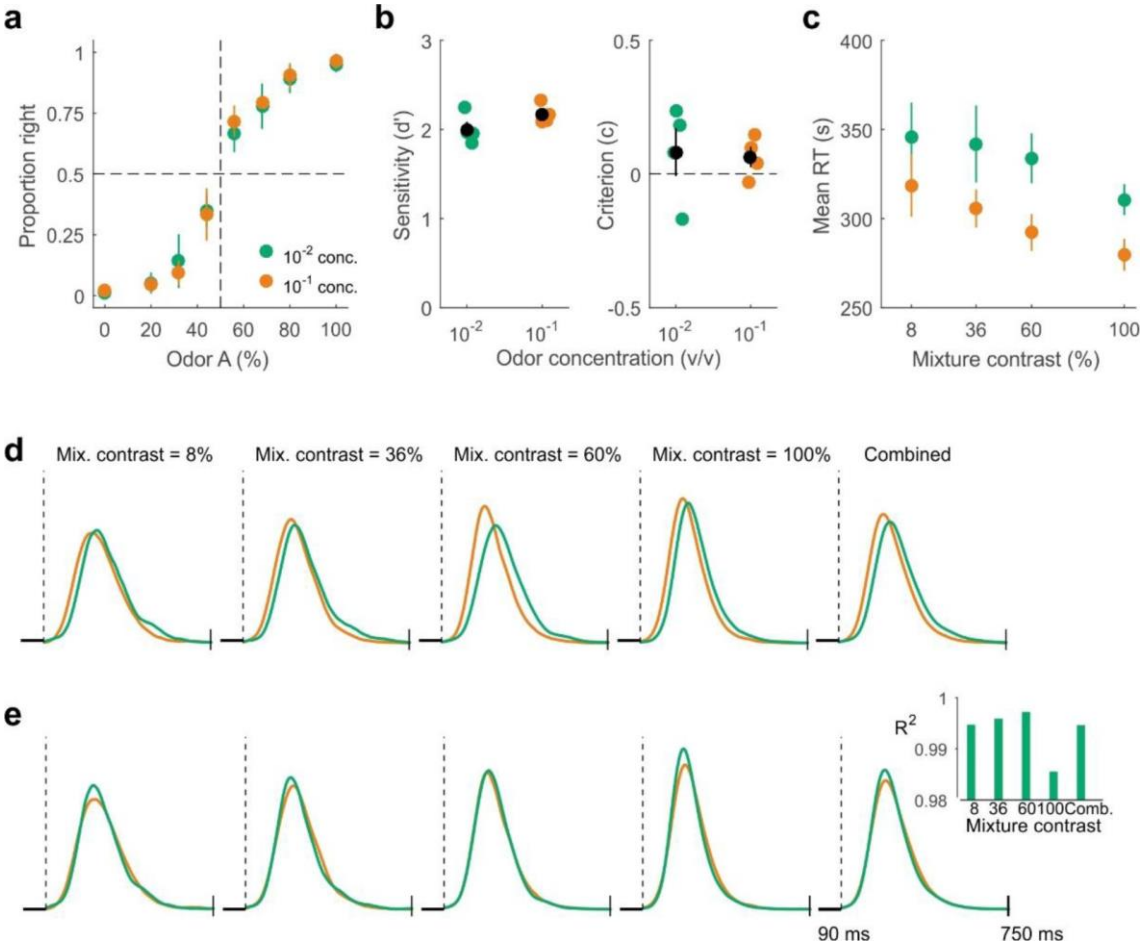
(gray) or ILD (black) on the RTDs from (a) and (b). Lines show linear fits. The accuracy of the scale invariance of the RTDs induced by changes in ABL is independent of the magnitude of the slope (slope of the linear fit not significantly different from zero; $p = 0.6$, Permutation test) and thus of the difference between the two RTDs being compared. In contrast, for changes in difficulty, the difference in shape (not just scale) of the RTDs becomes larger the larger the ILD-induced change in RT, as evident by statistically significant negative slope of a linear fit of R^2 versus slope ($p = 0.003$, Permutation test), showing that R^2s in this case are only large because the speed accuracy trade-off in the task is weak. These results are in agreement with the claims made in the main text, whereby ABL-induced scale invariance is exact, whereas difficulty-induced scale invariance is approximate and degrades with the difference in difficulty between the two RTDs. For a thorough treatment of the nature and form of the changes in the RTD induced by difficulty, see Section 4 of the Supplementary Note.

Supplementary Figure 4 The TIED holds in human ILD discrimination.



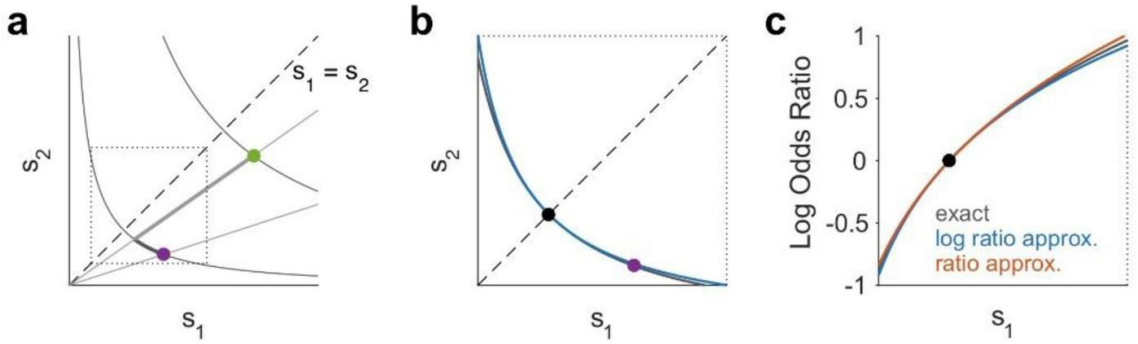
Same format as Fig. 2 of the main text. **(a)** Psychometric functions (n = 9 human subjects; circles, across-subject mean; lines, across-subjects standard deviation). **(b)** Discriminability index (left) and criterion (right) for each subject. **(c)** Chronometric functions ((n = 9; circles, across-subject mean; lines, across-subjects SEM). **(d)** RTDs for each ILD and for all ILDs combined. **(e)** Same but with green RTD rescaled (as explained in Methods, Supplementary Fig. 3) to maximize overlap with the orange RTD. **(inset)** Fraction of variance of the RTD at one ABL explained by the RTD at the other ABL (see Methods). In all panels, green (orange) corresponds to ABL = 40 (60) dB SPL. See Methods for details of the behavioral task.

Supplementary Figure 5 The TIED holds in rat olfactory mixture discrimination.



Same format as Fig. 2 of the main text. **(a)** Psychometric functions (n = 4 rats; circles, across-subject mean; lines, across-subjects standard deviation). **(b)** Discriminability index (left) and criterion (right) for each subject. **(c)** Chronometric functions (n = 4 rats; circles, across-subject mean; lines, across-subjects SEM). **(d)** RTDs for each mixture contrast and for all of them combined. **(e)** Same but with green RTD rescaled (as explained in Methods, Supplementary Fig. 3) to maximize overlap with the orange RTD. **(inset)** Fraction of variance of the RTD at one mixture contrast explained by the RTD at the other mixture contrast (see Methods). In all panels, green (orange) corresponds to an odor concentration of 10^{-2} (10^{-1}) (v/v). See (Mendonça, A., et al., *bioRxiv*, page 501858, 2019) for details of the behavioral task.

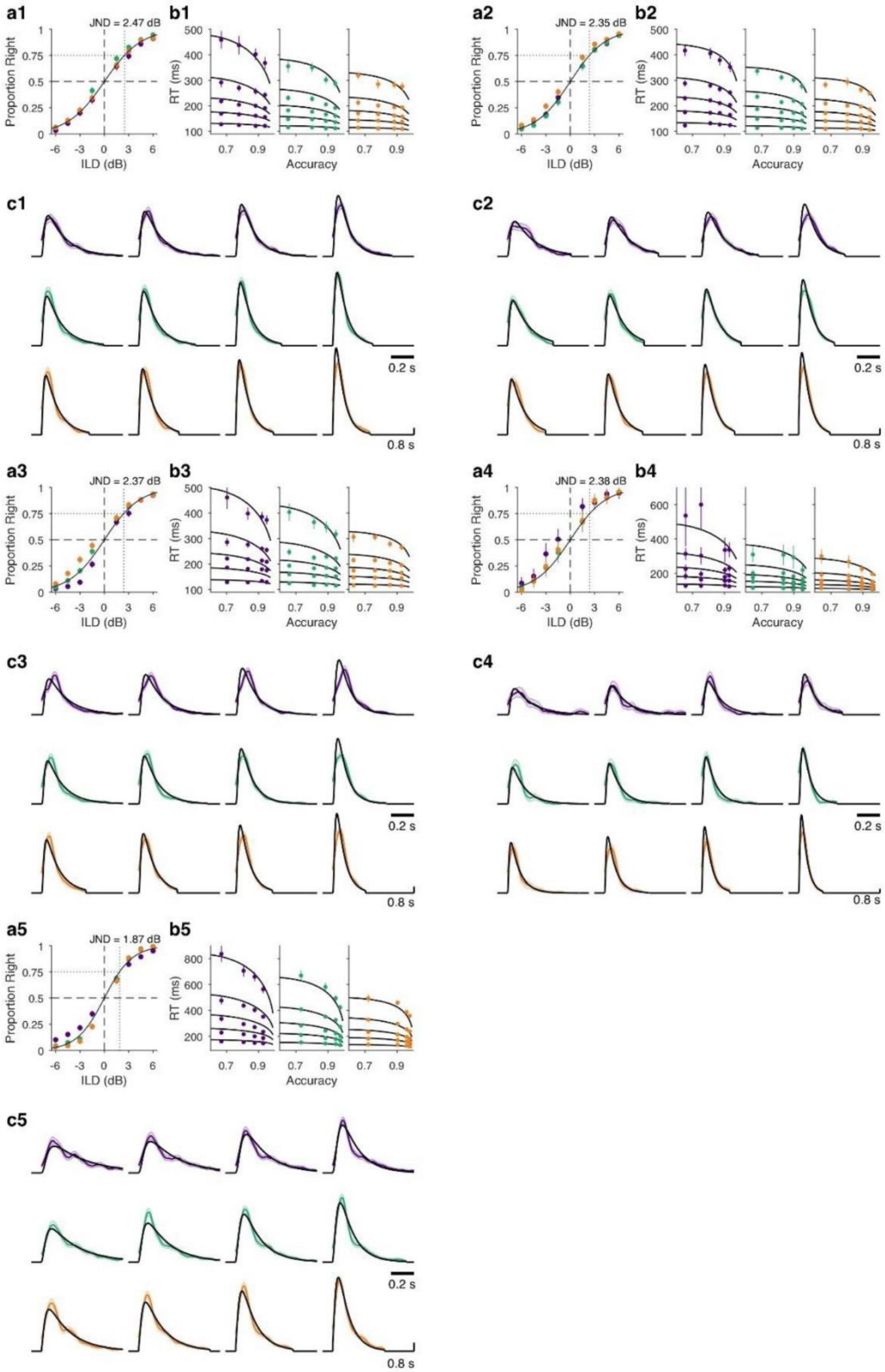
Supplementary Figure 6 Relationship to the standard drift-diffusion model (DDM) and relevance of the log-ratio.



(a) In the standard use of the DDM, only the drift coefficient μ is stimulus-dependent and the variance σ^2 of the evidence is constant (Ratcliff, R. and Rouder, J., *Psychological Science*, 9(5):347-356, 1998; Palmer, J., et al., *Journal of vision*, 5(5):1-1, 2005). In contrast, if the activity of the sensory neurons is Poisson, both the mean and the variance of the evidence will in general be stimulus-dependent. In the figure, two discrimination problems (filled circles) are shown in stimulus space (s_1, s_2). We consider the type of stimulus dependence which holds in our experiments, $\mu \propto s_1^\lambda - s_2^\lambda$ and $\sigma^2 = s_1^\lambda + s_2^\lambda$. The difference between any two problems can be decomposed into a difference in the difficulty of the problem (gray curved lines) while keeping the variance of the evidence constant, and a difference in the effective time units of the problem (straight dark-gray lines) while keeping the intensity-ratio constant. Moving along these straight lines corresponds to a uniform stretching of the RTD; neither choice-accuracy nor the shape of the RTD (only its scale) change along this line. Moving along the equi-variance lines we change the ratio of the stimulus intensities, and thus both choice accuracy and the shape of the RTD vary. Thus, when we speak of keeping the 'overall intensity' of the stimuli constant in the main text, we are formally referring to stimulus manipulations which keep the variance of the evidence constant. **(b)** In general, keeping the variance of the evidence constant while the stimuli change requires a good understanding of how the evidence relates to the stimulus. In our simple model, it requires knowing the value of λ. However, close to psychophysical threshold, keeping the variance constant is easier, since the constant σ^2 line becomes perpendicular to the identity line s_1=s_2. Thus, any changes which keep constant any symmetric function of the stimulus are expected to also keep the variance approximately constant. This panel shows a zoomed version of the dashed square region in (a), representing discrimination problems close to psychophysical threshold (i.e., $s_1 \sim s_2$). Gray line is as in (a), and blue line is the line representing constant ABL which we use in our experiments. Keeping $s_1 + s_2$ constant would be a worse approximation, but which would nevertheless be appropriate if s_1 and s_2 are sufficiently similar. **(c)** Here we considered whether there is anything special about the logarithmic transform. In a signal detection theory model with additive noise and a logarithmic encoding of the stimulus intensity (Fechner's model (Fechner, G., Breitkopf and Harterl, 1860)), the d' of a discrimination is proportional to

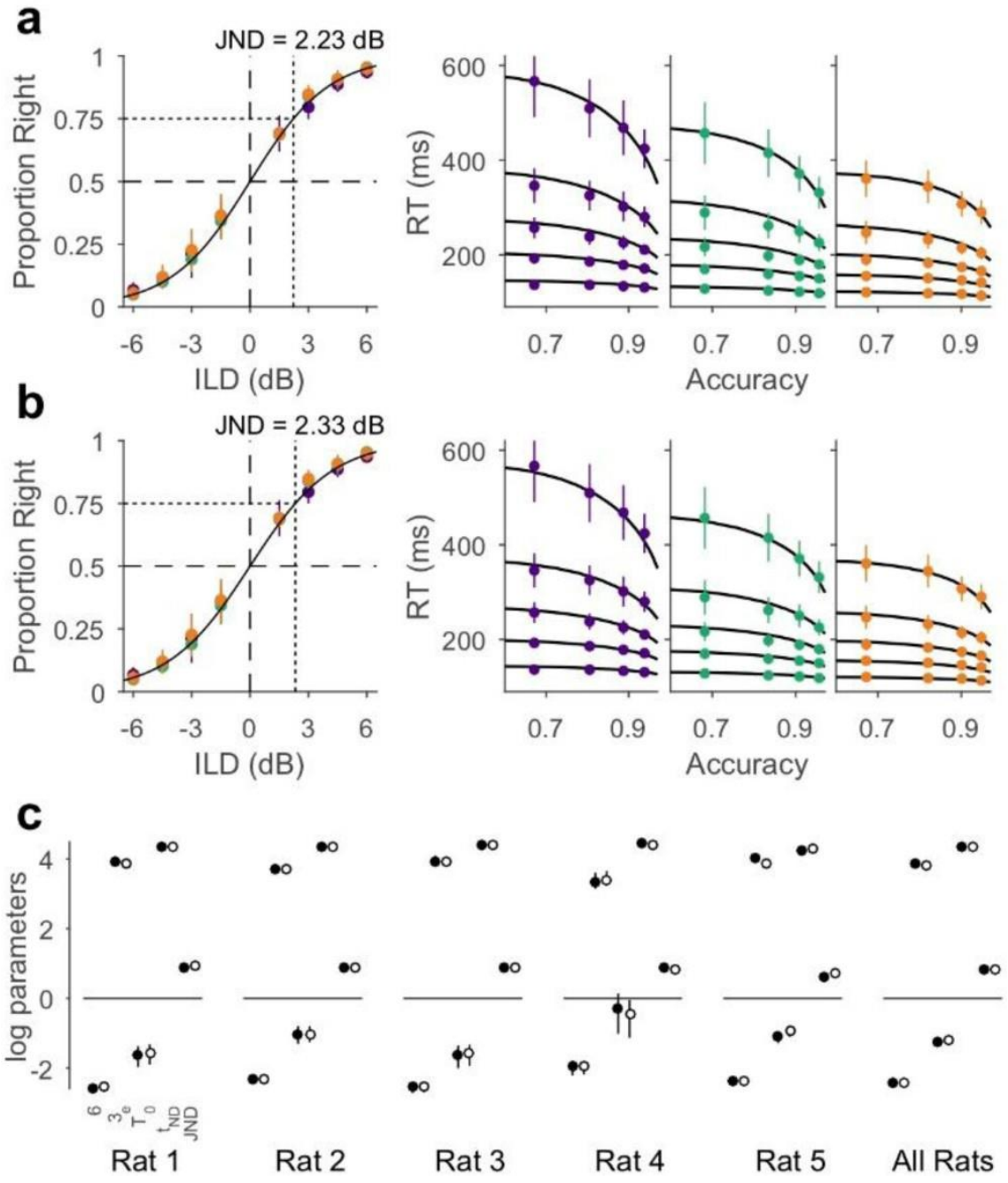
the log of the ratio of the stimulus intensities. We also obtain that the choice log-odd ratio (LOR) close to psychophysical threshold is proportional to ILD (Eq.44), i.e., the log of the ratio of the RMS pressure level of the two stimuli (Fig. 1c). However, we have not explicitly invoked a logarithmic transformation. Logarithms only appear in our description because sound level is typically measured in dB. In deriving Eq. 44, which is valid near threshold, we took the exact expression for the choice-LOR (Eq. 38) and developed it to first order in ILD. However, one can also develop this expression to first order in the small quantity $1-s_1/s_2$. The figures shows the choice log-odds ratio (LOR) as one varies one stimulus (s_1) while moving along the gray (gray line) and blue (blue and red) lines in (b). Gray is the exact LOR. Blue is the linear approximation in $\log s_1/s_2$, equivalent to Eq. 44. Red is a linear approximation in $(s_2-s_1)/s_2$. These two approximations are similarly accurate, at least for the range of difficulties that we have used. Thus, although the the mathematical description of our model is compact when measuring sound intensity in dB, there is nothing unique about the logarithmic transformation. For all panels in this figure, we used λ=0.1.

Supplementary Figure 7 Model fits for individual rats.



(a1-5) Psychometric functions. Dots are choose-right probabilities separately for each ABL (same color code as Fig. 4 main text). A single psychometric function has been fit to pooled data across ABLs (mean ± SD for each individual rat). **(b1-5)** The five quantiles for the RTDs for the three ABLs and their model fit. Error bars are 95% CIs across bootstrap resamples (N_r = 1000). **(c1-5)** RTDs for each rat and their corresponding model fit across all twelve conditions. Shaded regions represent 95% CIs across bootstrap resamples (N_r = 1000). Results are more noisy than for the pooled data but the model can still fit the behavior of each single rat accurately.

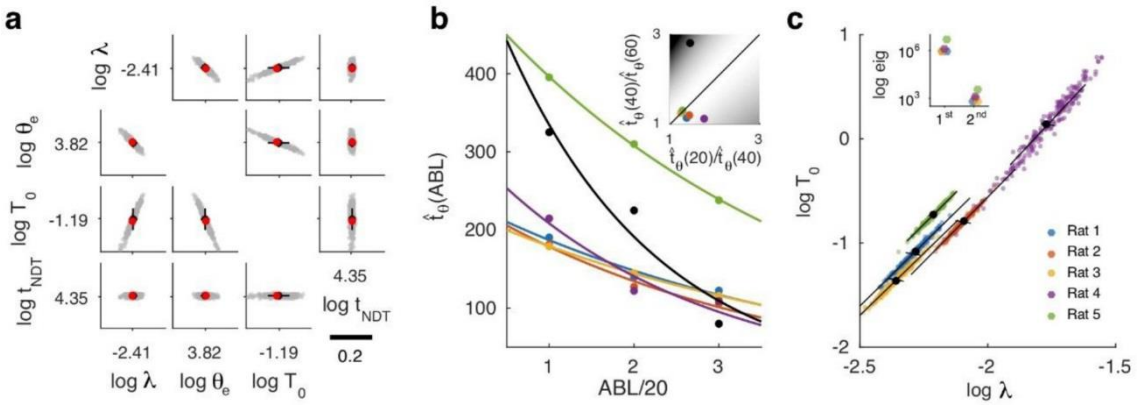
Supplementary Figure 8 Constrained versus Unconstrained model comparison.



(a) Constrained fitting approach. In the main text (Fig. 4) we used a 'constrained' model fitting approach. For this approach, we first fit $\Gamma \sim \lambda\theta_e$ from the psychometric function of the rats. Then, assuming that Γ is now fixed, we fit the remaining three parameters λ, T_0, t_{NDT} using only the two extreme values of ABL = 20 and 60 dB SPL (see Methods). The figure shows the results of this fit, which is the same data shown in Fig. 4b, c (n = 5 rats). Left: Choose-right probabilities (circles, mean ± across rats, n = 5; black line,

model fit). Right: circles show the 0.1, 0.3, 0.5, 0.7 and 0.9 quantiles (mean ± SEM across rats, n = 5, black line, model fit). The variance in the shape of the observed RTDs explained by the constrained model is 97.7 ± 2.3% (see Methods, main text). **(b)** Unconstrained fitting approach. Alternatively, one can fit all four parameters simultaneously from all the data (RT and accuracy from all conditions, see Methods). The figure shows the results of this fitting approach in the same format as in (a). The variance in the shape of the observed RTDs explained by the unconstrained model is 98.4 ± 1.4% **(c)** Log parameter estimates (and JND) for the constrained (black circles) and unconstrained (white circles) models for the pooled data and for individual rats (n = 5 rats, error bar represents 95% bootstrap CI). The model fits and estimated parameter values using both model fitting approaches are almost identical. The fact that an unconstrained fit produces the same result as the constrained fit means that both the coupling between speed and accuracy as a function of ILD, as well as the effect of ABL on RTs, have the same nature in the model and in the data. Supplementary Table 3 also shows the values of the estimated parameters for both model fitting approaches using the Quantile Maximum Likelihood method (Heathcote, A. and Brown, S., *Psychonomic Bulletin and Review*, 11(3):577-578, 2004) (Methods), which are effectively identical to those obtained using the χ^2 Methods.

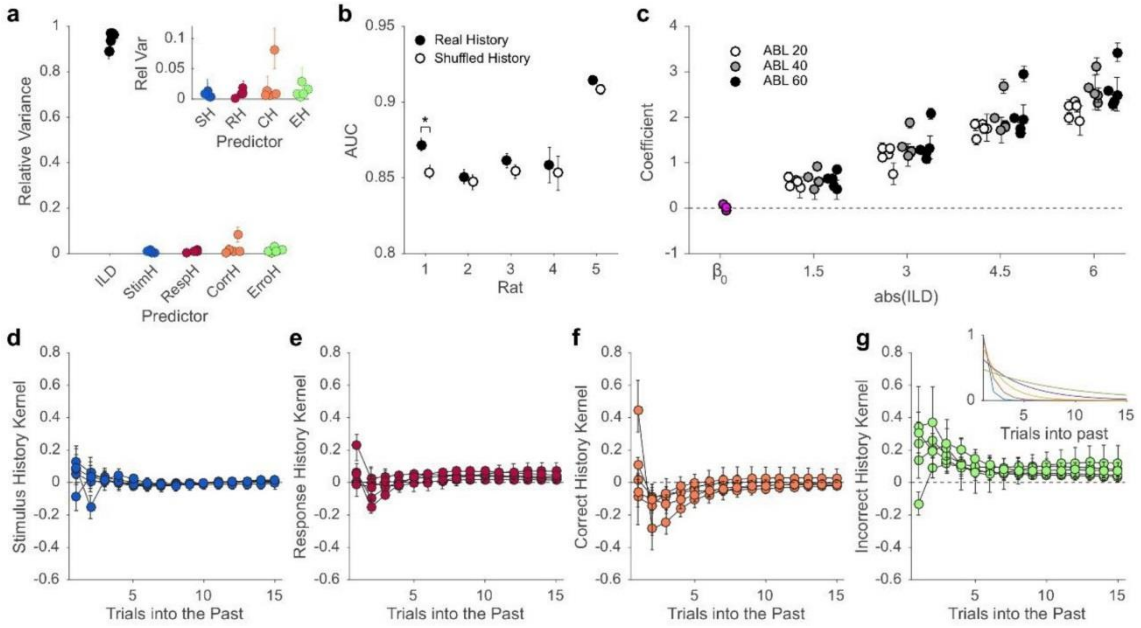
Supplementary Figure 9 Uncertainty in parameter estimation**.**



(a) We obtained an estimate of the joint probability distribution of the parameters from (unconstrained) model fits of bootstrap re-samples of our dataset. Gray: Log-parameters from 200 re-samples. Black: mean ± standard deviation across 1000 re-samples (n = 5 rats). Red dot is the fit to the actual data. There are strong correlations between λ, θ_e and T_0. The negative correlation between λ and θ_e is easy to explain, since accuracy only constrains their product. The figure shows that this product (or, equivalently Γ, which sets the rat's JND) is very well specified by the data. **(b)** To gain an understanding of the relationship between T_0 and λ we explored how they jointly determine the dependency of RT on ABL. Recall that we have assumed that $RT=t_{NDT}+\tau_{DT}t_\theta$ and that this temporal rescaling relationship describes our data very accurately (Fig. 4). In order to obtain a model-independent estimate of t_θ, we used a procedure analogous to the one in Supplementary Fig. 3, and inferred this estimate t^θt^θ, linearly regressing the quantiles of the actual RT distributions for each value of ABL on those for the ABL-independent distribution of τ_{DT} specified by the dimensionless DDM in Eq. 41. We performed a single fit for all difficulties with a given ABL. Thus, we have $\hat{t}_\theta(ABL) = (\chi^2 T_0 \Gamma^2/2\lambda^2)10^{-\lambda ABL/20}$ which we view as a non-linear equation with ABL as the independent variable for which we have three data points per rat. Since Γ is very well specified by the data, we assume it is constant and equal to its best-fit value and consider T_0 and λ as the parameters of interest. We used non-linear least squares to study how well λ and T_0 are specified by this last equation. The fits were actually performed on log-parameters to avoid ambiguities due to units of measurement. The figure shows the empirical time-scale factor t^θ(ABL)t^θ(ABL) for each of the three ABL conditions and the model fits. Each color except black is for an individual rat. **(c)** Estimates of log T_0 versus log λ from the non-linear least square fit. Colored points are estimates from bootstrap re-samples. For each color, black full circle is the mean across 1000 re-samples. Black lines represent the directions of the eigenvectors of the Fisher Information Matrix (FIM) evaluated at the best fit, with lengths proportional to the covariance of the parameters (they are not orthogonal because the x- and y-axis are stretched). **Inset**. The two eigenvalues of the FIM for each rat. The eigenvalues span several orders of magnitude, a signature of that the there is a 'sloppy' direction in parameter space in the model

(the one along which log λ and log T_0 are positively correlated) (Transtrum, M., et al., *Physical Review E*, 83(3):036701, 2011; Machta, B., et al., *Science*, 342(6158):604-607, 2013). Inspecting the functional form of $\hat{t}_\theta(ABL)$, we see that its curvature with respect to ABL is determined by λ, and that λ and T_0 jointly determine the overall range of $\hat{t}_\theta(ABL)$. Sloppiness arises because similar curves can be produced by small correlated changes in curvature and range. Although curvature and range also appear correlated across rats, we believe that more rats and more values of ABL are needed to establish that this correlation is indeed a robust feature of the data. Despite sloppiness, the model is still quite sensitive to the values of $\hat{t}_\theta(ABL)$. For instance, it is straightforward to check that the following equality should hold $\hat{t}_\theta(ABL = 40)/\hat{t}_\theta(ABL = 60) = \hat{t}_\theta(ABL = 20)/\hat{t}_\theta(ABL = 40)$. Artificial values of $\hat{t}_\theta(ABL)$ perfectly within the range of what we observed, but which strongly violate this equality (black circles in panel b), give poor model fits (panel b, inset).

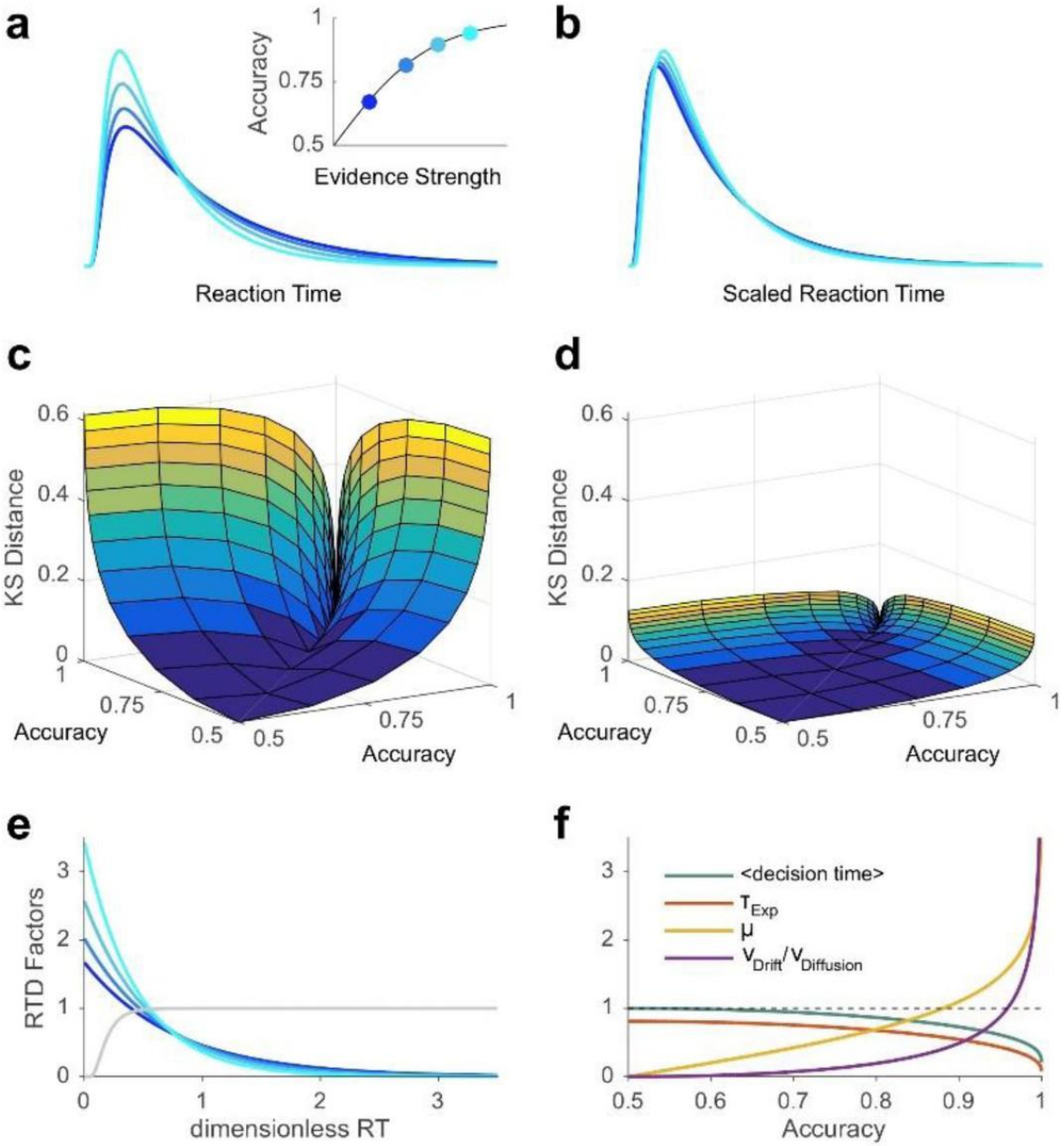
Supplementary Figure 10 Trial history effects.



In order to reveal the extent to which variables different from the sensory stimulus in the current trial had control over the rat's responses (Busse, L., et al., *Journal of Neuroscience*, 31(31):11351-11361, 2011; Fründ, I., et al., *Journal of vision*, 14(7):9-9, 2014), we quantified the predictive power of the history of stimuli, responses, responses after correct outcomes and responses after errors on choices in the current trial using logistic regression (n = 5 rats, see Methods). (**a**, **b**) Same data as in Fig. 5c. (a) Fraction of variance of the linear component of the logistic regression model captured by the stimulus and the four types of history effects we considered. Each dot is a single rat. Inset. Zoom in on the fraction of variance associated to the history effect predictors. (**b**) Area under the curve (AUC) for each rat for the actual data and for surrogate datasets where history predictors have been shuffled (see Methods). (**c**) Current trial stimulus predictor coefficients separately for each ILD and ABL. Coefficients are mostly ABL-independent and grow linearly with ILD. The linear increase with ILD is a signature of discriminations at psychophysical threshold (see Section 3.3 of the Supplementary Note). (**d**–g) Kernels for each of the four types of history predictors. Notice the difference in the scale of the y-axis between this panel and panel (c). Dots connected by lines indicate the kernel for each single rat. The larger positive-then-negative kernel in (f) (thick-line) corresponds to Rat 1, which shows a small but significant effect of history (point with relative variance > 5% in panel a, inset). This rat had a mild win-stay strategy due to a tendency to leave the back of her body aligned with the rewarded lateral port for the subsequent trial. Inset in (g) shows the basis of decaying exponentials used to express all kernels. Across the figure, error bars are bootstrap 95% CI (2000 re-samples, see Methods). These results suggest that stimulus-independent

strategies have only marginal control over the responses of the rats in the task, and that their choices on a trial-by-trial basis reflect almost exclusively their perception of the ILD of the sounds in the current trial.

Supplementary Figure 11 RT scale invariance with respect to changes in strength of evidence.



The text and mathematical details associated to this figure are in section 4 of the Supplementary Information. **(a)** Four representative RTDs from the drift-diffusion model. **Inset**. Accuracy as a function of strength of evidence. **(b)** Same distributions but in rescaled time. The time axis for the three slowest distributions was changed so as to maximize the overlap (minimize the Kolmogorov Smirnoff (KS) distance) with the fastest one. **(c)** KS distance as a function of the accuracy associated to any pair of RT distributions. **(d)** Same after temporal rescaling. **(e)** Exponential and refractory components of the RTDs shown in (a), with the same color code (see section 4 of the Supplementary Note). The refractory term

(gray) is the same for all of them as it does not depend on the strength of evidence. The color lines show $K(\mu, \sigma^2)E(t)$ (see Eq. 53, 54), so that the product of the two lines is the actual RTD. **(f)** Properties of the exponential decay term E(t) as a function of accuracy. First two lines are the mean decision-time and the time constant of the exponential decay term τ_{Exp}. Third line is the strength of evidence. Fourth line is the ratio $v_{Drift}/v_{Diffusion}$ (see Eq. 50) Only for accuracies very close to unity does the strength of evidence come to dominate the shape of E(t) and thus of the RTD. The different quantities in this plot are all dimensionless and have comparable values; thus we've used a single y axis whose label should be inferred from the legend. For all plots in this figure we've assumed without loss of generality that $\sigma^2=1$.