frontiers in Psychology

ORIGINAL RESEARCH
published: 12 September 2018
doi: 10.3389/fpsyg.2018.01691

Context Effects in the Judgment of Visual Relative-Frequency: Trial-by-Trial Adaptation and Non-linear Sequential Effect

Xiangjuan Ren[1,2], *Muzhi Wang*[3] *and Hang Zhang*[2,3,4]*

[1] Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China, [2] Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China, [3] School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China, [4] PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

Humans' judgment of relative-frequency, similar to their use of probability in decision-making, is often distorted as an inverted-*S*-shape curve—small relative-frequency overestimated and large relative-frequency underestimated. Here we investigated how the judgment of relative-frequency, despite its natural reference points (0 and 1) and stereotyped distortion, may adapt to the environmental statistics. The task was to report the relative-frequency of black (or white) dots in a visual array of black and white dots. We found that participants' judgment was distorted in the typical inverted-*S*-shape, but the distortion curve was influenced by both the central tendency and spread of the distribution of objective relative-frequencies: the lower the central tendency, the higher the overall judgment (contrast effect); the higher the spread, the more curved the inverted-*S*-shape (curvature effect). These context effects are in the spirit of efficient coding but opposite to what would be predicted by Bayesian inference. We further modeled the context effects on the level of individual trials, through which we found not only a trial-by-trial adaptation, but also the non-linear sequential effects that were recently reported mainly in circularly distributed visual stimuli.

Keywords: probability distortion, subjective probability, frequency estimation, sequential effect, adaptation, Bayesian inference, efficient coding

OPEN ACCESS

Edited by:
Andrey R. Nikolaev,
KU Leuven, Belgium

Reviewed by:
Guido Marco Cicchini,
Consiglio Nazionale Delle Ricerche (CNR), Italy
Ambarish Pawar,
Salk Institute for Biological Studies, United States
Ingo Fründ,
York University, Canada

***Correspondence:**
Hang Zhang
hang.zhang@pku.edu.cn

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *10 April 2018*
Accepted: *22 August 2018*
Published: *12 September 2018*

Citation:
Ren X, Wang M and Zhang H (2018) Context Effects in the Judgment of Visual Relative-Frequency: Trial-by-Trial Adaptation and Non-linear Sequential Effect. Front. Psychol. 9:1691. doi: 10.3389/fpsyg.2018.01691

INTRODUCTION

The human perceptual system adapts to the environmental statistics from time to time (Helson, 1947; Gilchrist et al., 1999; Dean et al., 2005; Chopin and Mamassian, 2012; Gepshtein et al., 2013). For example, a lighted outdoor sign that dazzles at night may look dim in the daylight. Adaptation like this allows the human brain to use neurons of limited dynamic range to represent the immense dynamic range of physical stimuli (10^9 for luminance, from star light illumination to intense daylight conditions). But it comes at a cost: in order to be sensitive to differences in the current environment, the mapping from physical stimuli to perception must be non-stationary. That is, a stimulus that is physically 5 times as large as a second stimulus may be perceived 10 times as large as the latter in one context and only 2 times as large in a different context. This non-stationarity can be harmless in many situations (e.g., in the perception of lightness), where only ordering information (e.g., which is brighter and which is darker) is required.

Here we investigated how the judgment of relative-frequency may adapt to the environment. For the perception of relative-frequency, adaptation can be both helpful and harmful. On one hand, relative-frequency in real life, like luminance, has a vast dynamic range that may challenge the neural system. For example, the relative-frequencies of different causes of death span six orders of magnitude (Lichtenstein et al., 1978). On the other hand, as a source of probability information, relative-frequency needs an accurate representation. Any non-stationary transformations accompanying adaptation would hurt one's ability to maximize expected gain in decision-making.

Relative-frequency differs from many sensory stimuli in its abstractness and in its finite range—from 0 to 1. What is special about relative-frequency is also its stereotyped distortion: Humans' judgment of relative-frequency, similar to their use of probability in decision-making (Tversky and Kahneman, 1992; Gonzalez and Wu, 1999), is often distorted in an inverted-S-shape—small relative-frequency overestimated and large relative-frequency underestimated. For example, people overestimate the relative-frequency of rare causes of death such as flood and hurricane and underestimate that of common causes such as heart disease (Lichtenstein et al., 1978; see Zhang and Maloney, 2012 for more examples). The opposite pattern, S-shaped distortion, was also reported (Shuford, 1961; Pitz, 1966; Brooke and MacRae, 1977; Wu et al., 2009). Zhang and Maloney (2012) found that the inverted-S- or S-shaped distortion in a variety of tasks could be well-captured by a Linear-in-Log-Odds (LLO) transformation:

$$\lambda\left[\pi\left(p\right)\right] = \gamma\lambda\left[p\right] + (1-\gamma)\lambda\left[p_0\right], \quad (1)$$

where p and $\pi\left(p\right)$ respectively denote objective and subjective probability or relative-frequency, $\lambda[\cdot]$ denotes the log-odds transformation, $\lambda\left[p\right] = \log\frac{p}{1-p}$, and γ and p_0 are free parameters that are readily interpretable. The parameter γ indicates the slope of the distortion curve, with $\gamma < 1$ for inverted-S-shaped distortion, $\gamma = 1$ for no distortion, and $\gamma > 1$ for S-shaped distortion. The parameter p_0 indicates the crossover point where $\pi\left(p\right) = p$. In other words, the γ and p_0 are measures respectively for the curvature and elevation of probability distortion (Gonzalez and Wu, 1999).

In our experiment, participants judged the relative-frequency of black (or white) dots among an array of black and white dots (Figure 1A). There were four conditions for the distribution of the objective relative-frequency p (Figure 1B). In the baseline Uniform condition, p was uniformly distributed between 0.01 and 0.99. The Small (Large) condition differed from the Uniform condition mainly in the central tendency of the distribution by having a disproportionally great number of very small (large) values of p. The Extreme condition was U-shaped (i.e., most values were extreme) and differed from the Uniform condition in the spread of the distribution.

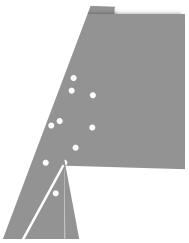
We asked two questions. The first question is whether and how participants' judgment of relative-frequency, $\pi\left(p\right)$, may vary with the distribution of p. There has been increasing evidence that adaption functions not only for sensory modalities, but also for abstract quantities such as utility (Tobler et al., 2005; Kobayashi et al., 2010; Louie et al., 2013; Khaw et al., 2017; Rustichini et al., 2017), numerosity (Burr and Ross, 2008; Cicchini et al., 2014), rate (Levitan et al., 2015), and variance (Payzan-LeNestour et al., 2016), in the form of contrast effects: the same quantity tends to be perceived larger in a context of small quantities and smaller in contrast with large quantities. Such contrast effect was also found for relative-frequency in a task similar to ours (Varey et al., 1990).

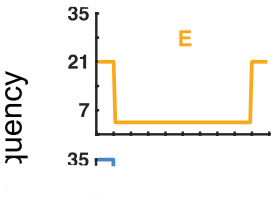
What concerned us are not individual values but how the whole curve of $\pi\left(p\right)$ may change with the context and what principles the changes may follow. We considered two lines of theories that provide opposite predictions for the possible context effects (see Figure 2 for the simulated predictions). One line of theories is represented by the adaption-level theory (Helson, 1947; Parducci, 1965), which assumes that the perception of a specific stimulus reflects the difference between the stimulus and an internal reference point. The value of the reference point, called "adaptation level", is determined by the average value of the stimuli in the context. The adaptation-level theory predicts a contrast effect (Figure 2A): The Small condition, which had a lower central tendency than the Uniform condition, would lead to a higher elevation for the $\pi\left(p\right)$ curve, while the Large condition would lead to a lower elevation than the Uniform condition. Because the adaptation level is not influenced by the spread of the distribution, the adaptation-level theory predicts no difference between the Extreme condition and the Uniform condition.

The second line of theories treats perceptual judgment as a Bayesian inference problem [see (Maloney and Zhang, 2010; Petzschner et al., 2015) for reviews]—inferring the true value of a physical stimulus (in the current experiment, relative-frequency) based on its noisy percept. To compensate for the uncertainty in the percept, the final judgment would combine the percept and prior information about the stimulus. If the prior participants used follows the distribution of p's they had experienced in the experiment, their judgment would be biased toward the high-density regions of the distribution. Thus, the Bayesian inference theory predicts an assimilation effect (Figure 2B): The Small condition, which had high densities on the small end, would have a lower elevation than the Uniform condition, while the Large condition would have a larger elevation than the Uniform condition. Similarly, for the U-shaped Extreme condition, the concentration of p's on the two ends would attract $\pi\left(p\right)$toward the two ends, that is, a steeper slope than the Uniform condition.

Our second question is how the context effects, if any, may arise from trial to trial. In our experiment, participants were never explicitly informed about the distribution of p and could only learn the distribution via individual trials. We modeled two processes on the level of individual trials. The first process is a trial-by-trial updating of reference point, which is a natural extension of the adaptation-level theory with the additional assumption that the adaptation level (reference point) is updated by the delta rule (Rescorla and Wagner, 1972). As the result, the reference point assigns higher weights to more recent trials and would be able to track the changes in the context.

The second process we investigated is the sequential effect, that how the stimulus or response of a precedent trial may




35
21
7
E
quency

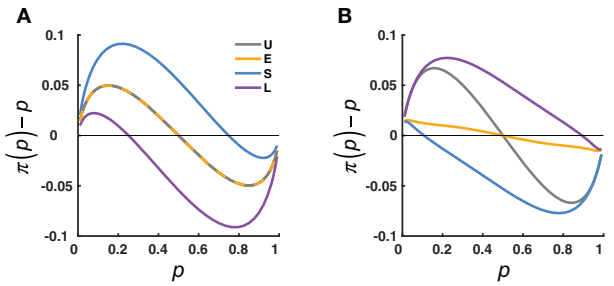


FIGURE 2 | Opposing effects predicted by two influential lines of theories**.** The predicted deviation of the subjective from objective relative-frequency, $\pi(p) - p$, is plotted as a function of the objective relative-frequency, p, and compared across the four distribution conditions (U, E, S, L, color coded). **(A)** Adaptation-level theory (Helson, 1947). The $\pi(p)$ is assumed to reflect the

distribution as their estimate for p. The estimate conditional on a specific percept y is

$$\pi_y = \sum_q q \, \mathbf{Pr}(q \mid y). \tag{6}$$

Given that the percept y itself is a random variable that cannot be directly observed, we need to marginalize o y to obtain a mapping from p to the final estimate $\pi(p)$:

$$\pi(p) = \int \pi_y \, \mathbf{Pr}(y \mid p) \, dy. \tag{7}$$

For Figure 2B, the parameter $\sigma_{noise} = 1$.

A aptat on L v ory

The adaptation-level theory does not predict the inverted-S-shaped distortion itself but predict how the distortion may change with the context. In our simulation for the adaptation-level theory, the $\pi(p)$ is determined by the same equation as LLO except for the inclusion of the adaptation level L:

$$\lambda[\pi(p)] = \gamma(\lambda[p] - \lambda[L]) + (1-\gamma)\lambda[p_0], \tag{8}$$

where, as in LLO, γ and p_0 are free parameters, and $\lambda[\cdot]$ denotes the log-odds transformation. The value of L shifts with the distribution of p:

$$\lambda[L] = \eta \sum_p \theta(p) \lambda[p], \tag{9}$$

where η is a free parameter and $\theta(p)$ is defined for each distribution condition as in Equation (4).

For Figure 2A, the parameters $\gamma = 0.8$, $p_0 = 0.5$, $\eta = 0.2$.

Measures of Distortion of Relative-Frequency

op an Crossov r o nt Est at Fro LL

For each participant, we used LLO (Equation 1) to fit the reported relative-frequency, $\pi(p)$ and estimated the slope parameter γ and the crossover point parameter p_0.

oot D stort on Curv an on_para tr c M asur s

The γ and p_0 provide a model-based summary for a distortion curve. Still, critical details of the curve may be lost due to the limitation of the model. As a complementary analysis, we smoothed the distortion curve for each participant and elicited non-parametric measures of distortion from the smoothed curve.

In particular, we smoothed $\pi(p) - p$ using a kernel regression method with the commonly-used Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964; Aljuhani and Al turk, 2014):

$$\hat{M}_h(x) = \frac{\sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right)}, \tag{10}$$

where

where σ_k denotes the span of the Gaussian kernel and was set to be 0.1, p_j and p_{j-1} respectively denote the objective relative-frequency of trial j and trial $j-1$, $j = 2, 3, ..., N$. If we define

$$W = \begin{pmatrix} w_1^2(p_n, p_{n-1}) & 0 & \cdots & 0 \\ 0 & w_2^2(p_n, p_{n-1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N^2(p_n, p_{n-1}) \end{pmatrix}, \quad (14)$$

$$X = \begin{pmatrix} S_2 & R_1 & 1 \\ S_3 & R_2 & 1 \\ \vdots & \vdots & \vdots \\ S_N & R_{N-1} & 1 \end{pmatrix}, \quad (15)$$

$$Y = \begin{pmatrix} R_2 \\ R_3 \\ \vdots \\ R_N \end{pmatrix}, \quad (16)$$

the coefficients of the weighted least-square regression at (p_n, p_{n-1}) could be estimated as:

$$\begin{pmatrix} \hat{\beta}_0^{WLS} \\ \hat{\beta}_{-1}^{WLS} \\ \hat{\beta}_C^{WLS} \end{pmatrix} = \left(X^T W X\right)^{-1} X^T W Y. \quad (17)$$

Modeling

We considered six models of $\pi(p)$, which are all based on LLO but differ in two dimensions: whether to include trial-by-trial adaptation and the type of sequential effects assumed. In the equations for all models, R_n (= $\lambda\left[\pi(p_n)\right]$ = log $\frac{\pi(p_n)}{}$

TABLE 1 | Notations.

VARIABLES OR FUNCTIONS	
p	A generic value on the probability scale; objective probability or relative-frequency
$\boldsymbol{\lambda}(\cdot)$	Log-odds function of probability or relative-frequency . $\lambda(p) = \log(p/(1-p))$
$\pi(p)$	Subjective probability or relative-frequency
p_n	Objective probability or relative-frequency of Trial n
S_n	Stimulus of Trial n in log-odds. $S_n = \lambda(p_n) = \log(p_n/(1-p_n))$
R_n	Response of Trial n in log-odds. $R_n = \lambda(\pi[p_n]) = \log(\pi[p_n]/(1-\pi[p_n]))$
MODEL ABBREVIATIONS	
LLO	Linear in log-odds model
AL	Adaptation-level model
LLO-L	Linear in log-odds model with linear sequential effects
LLO-NL	Linear in log-odds model with non-linear sequential effects
AL-L	Adaptation-level model with linear sequential effects
AL-NL	Adaptation-level model with non-linear sequential effects
MODEL PARAMETERS	
γ	Slope of the linear transformation of log-odds
p_0	Crossover point; controlling the intercept of the linear transformation of log-odds
β_0	Coefficient for the S_n term
β_{-i}	Coefficient for the R_{n-i} term, $i = 1,2,...,5$
β_C	Coefficient for the constant term
σ_{noise}	

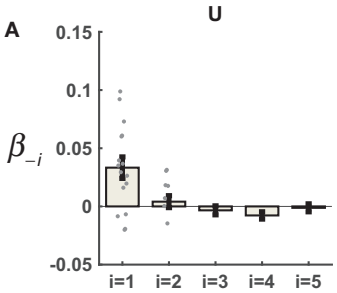

A
U
0.15
0.1
0.05
0
-0.05
β_{-i}
i=1
i=2
i=3
i=4
i=5

TABLE 2 | Pearson's *r* between data and model predictions in the pattern of sequential effects.

Condition	LLO	AL	LLO-L	AL-L	LLO-NL	AL-NL
Uniform	−0.015	−0.048	−0.215	−0.217	0.773	0.668
Extreme	−0.355	−0.174	−0.126	−0.318	0.723	0.588
Small	−0.141	0.219	−0.367	0.120	0.709	0.666
Large	0.238	−0.046	0.137	0.041	0.577	0.657

Mo[J] _Co, par son

To compensate for the di□erence in number of parameters between models, we computed the Akaike information criterion corrected for small sample-size (AICc; Akaike, 1974; Hurvich and Tsai, 1989),

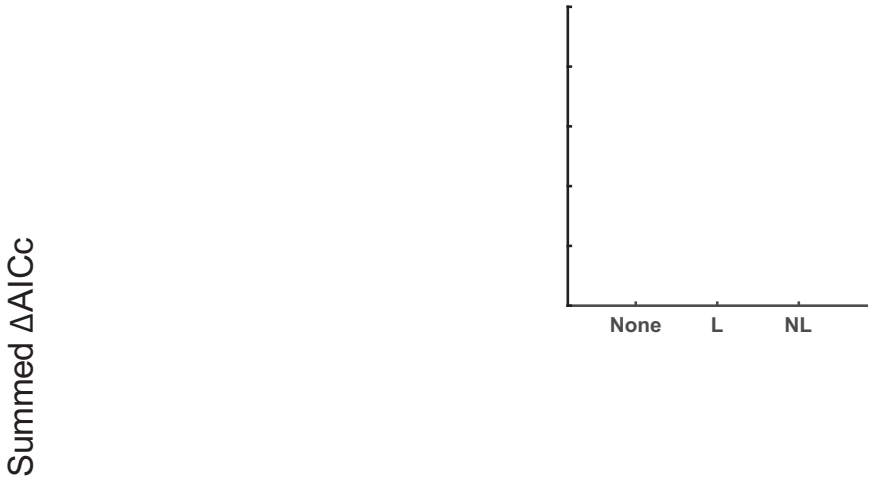
$$AICc = -2\ln\left(\hat{L}\right) + 2k + \frac{2k\left(k-1\right)}{N-k-1}, \quad (25)$$

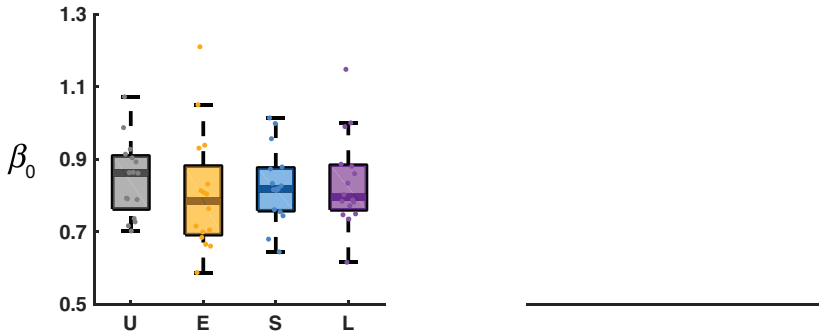
for each participant and each model as the metric for goodness-of-fit, where $\ln\left(\hat{L}\right)$ denotes the log likelihood maximized, k denotes the number of parameters, and N denotes the number of trials. The lower the AICc, the better the model fit.

The best model among the six models was the AL-NL model for all distribution conditions except for the Extreme condition (where the best was LLO-NL and the second best was AL-NL), according to the AICc summed across participants (Figure 5). A group-level Bayesian model selection (Stephan et al., 2009; Rigoux et al., 2014) based on AICc suggested the same (see the red dot in Figure 5 for the protected exceedance probability, that is, the probability a specific model is better than all the other models). We can also see that the models with non-linear sequential e□ects outperformed models with linear or none sequential e□ects, other things being the same. Except for the Extreme condition, AL models fit better than LLO models. The advantage of AL over LLO models was small in the Uniform condition and even negative in the Extreme condition, probably because the distribution of relative-frequency was centered at $p = 0.5$ in these conditions, where the final adaptation-level di□ered little from its initial value.

Est, at [J] ara, t rs

The estimated parameters for the AL-NL model are shown in Figure 6. The six parameters can be divided into four categories (see section Methods and Table 1 for more details): the slope and intercept parameters that belong to the original LLO model (β_0, β_C), the learning rate of adaptation-level (κ), the parameters that control the non-linear sequential e□ect (β_{-1}, ω), and the standard deviati909(l)0.05417046.8822044(i75]TJ-32.598-11.453(a)0Td[(T)909/R997g9.46497g925666-225.261-11.457Td[(p)0.059.4640.326)-0.24


Summed ΔAICc
None
L
NL


1.3
1.1
0.9
0.7
0.5
β_0
U
E
S
L

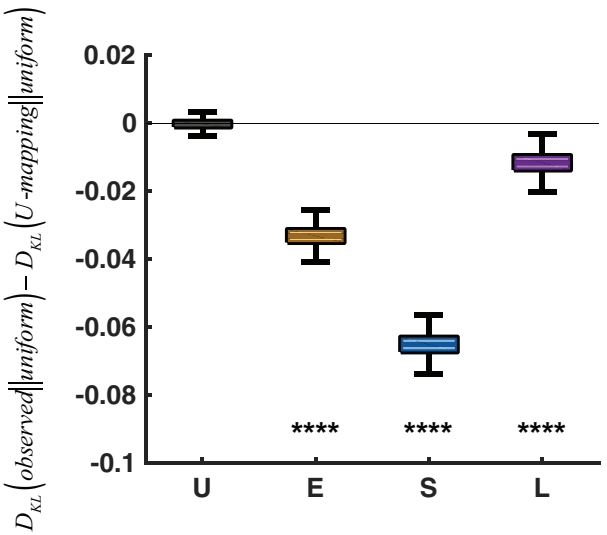


FIGURE 7 | Evidence for efficient coding. D_{KL} (*observed* ‖*uniform*) denotes the KL divergence from the uniform distribution over [0, 1] to the observed response distribution of a specific condition. D_{KL} (*U-mapping* ‖*uniform*) denotes the KL divergence from the uniform distribution to the response distribution predicted by the *p*-to-π (*p*) mapping of the Uniform condition. D_{KL} (*observed*‖ *uniform*) − D_{KL} (*U-mapping*‖ *uniform*) is plotted for each condition, with the Uniform condition serving as a sanity check (i.e., the difference should be 0). A negative difference implies that the *p*-to-π (*p*) mapping adopted by a specific condition is closer to efficient coding than applying the mapping of the Uniform condition to the condition. In the box plot, the middle line denotes the median across 1,000,000 simulations, the bottom and top lines denote the lower and upper quartiles, and the error bars denote the 99% confidence interval. ****

Bowman, A. W., and Azzalini, A. (1997). "Density estimation for inference," in *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-plus Illustrations,* **eds A. C. Atkinson, J. B. Copas, D. A. Pierce, M. J. Schervish, and D. M. Titterington (New York, NY: Clarendon Press, Oxford), 31.**

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* **10, 433–436. doi: 10.1163/156856897X00357**

Brooke, J. B., and MacRae, A. W. (1977). Error patterns in the judgment and production of numerical proportions. *Percept. Psychophys.* **21, 336–340. doi: 10.3758/BF03199483**

Burr, D., and Cicchini, G. M. (2014). Vision: e□cient adaptive coding. *Curr. Biol.* **24, 1096–1098. doi: 10.1016/j.cub.2014.10.002**

Burr, D., and Ross, J. (2008). A visual sense of number. *Curr. Biol.* **18, 425–428. doi: 10.1016/j.cub.2008.02.052**

Chopin, A., and Mamassian, P. (2012). Predictive properties of visual adaptation. *Curr. Biol.* **22, 622–626. doi: 10.1016/j.cub.2012.02.021**

Cicchini, G. M., Anobile, G., and Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc. Natl. Acad. Sci. U.S.A.* **111, 7867–7872. doi: 10.1073/pnas.1402785111**

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., and Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *J. Neurosci.* **32, 1056–1060. doi: 10.1523/JNEUROSCI.3411-11.2012**

Couzin, I. D. (2009). Collective cognition in animal groups.

Wei, X.-X., and Stocker, A. A. (2012). "E□cient coding provides a direct link between prior and likelihood in perceptual Bayesian inference," in *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems* **(Lake Tahoe).**

Wei, X.-X., and Stocker, A. A. (2015). A Bayesian observer model constrained by e□cient coding can explain 'anti-Bayesian' percepts. *Nat. Neurosci.* **18, 1509–1517. doi: 10.1038/nn.4105**

Wozny, D. R., Beierholm, U. R., and Shams, L. (2010). Probability matching as a computational strategy usded in perception. *PLoS Comput. Biol.* **6:e1000871. doi: 10.1371/journal.pcbi.1000871**

Wu, S.-W., Delgado, M. R., and Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proc. Natl. Acad. Sci. U.S.A.* **106, 6088–6093. doi: 10.1073/pnas.0900102106**

Zhang, H., and Maloney, L. T. (2012). Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. *Front. Neurosci.* **6:1. doi: 10.3389/fnins.2012.00001**

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.