The psychophysics of entropy judgments for musical note sequences

Hao Zhu

Department of Psychology, Brooklyn College of CUNY

Author Note

Correspondence concerning this article should be addressed to

Abstract

The study investigates the psychophysics of entropy judgments in the auditory domain, focusing on how individuals perceive entropy in musical note sequences. Building on Shannon's entropy model, the experiment aims to validate a procedure for measuring participant sensitivity to variations in entropy. Participants, recruited from Amazon's Mechanical Turk, listened to eight-second note sequences and assessed whether the notes were mostly the same or different. The entropy of pitches was systematically manipulated, ranging from fully random to highly predictable distributions. Feedback was provided to some participants to examine its effect on learning and adaptation. The results indicated that participants could discern different levels of entropy and placed their decision criterion around 2 bits. Various cognitive models were explored and simulated to interpret these findings. The study provides insights into the auditory perception of entropy, highlighting the potential for quantitative assessment of entropy sensitivity in the auditory domain.

Keywords: Entropy Judgment, Psychophysics, Auditory

The psychophysics of entropy judgments for musical note sequences

Introduction

Individuals have to confront the uncertainty inherent to their environment and future; as a result, perceiving and adapting toward information related to uncertainty is essential for survival. Shannon (1948) proposed a model of uncertainty that measured the amount of randomness in a discrete probability distribution as entropy. His entropy model has been applied across various sensory modalities to understand visual perception (Strange et al., 2005), auditory adaptation(De Pinho et al., 2002), and decision making (Cristín et al., 2022).

Building on Shannon's entropy model, our study narrowed its focus to statistical judgments of entropy in the auditory domain, particularly to how individuals perceive entropy in collections of musical note sequences. By exploring this specific aspect, we aimed to enhance the understanding of how auditory perception of entropy influences decision-making and whether this sensitivity can be quantitatively assessed. Thus, the experiment was designed to validate the general procedure and psychophysical approach for measuring participant sensitivity to variations in entropy among note sequences. Additionally, a range of possible psychological process models was built to explore the potential cognitive process of making these judgments.

The literature on statistical judgment provides a framework for understanding how people intuitively infer and make decisions based on statistical properties. For example, Peterson and Beach (1967) described how laypeople can function as 'intuitive statisticians,' capable of making judgments about statistical properties, such as proportions, mean, variance, even without formal statistical training. An experiment demonstrates how individuals estimate changing proportions by adapting their judgments based on variations in a sequence of two flashing lights (Peterson & Beach, 1967). This example shows that even in dynamic scenarios, individuals apply their understanding of statistical properties to predict outcomes. Essentially, they are not just passively receiving information but actively interpreting and reacting to it based on a statistical understanding of their environment.

Exploring whether humans and animals can become sensitive to the randomness of events,

or entropy, is valuable due to its potential implications for learning and adaptation in unpredictable environments. This capability is considered important for survival, as it helps species discern patterns and predictability in their surroundings. For example, a study on rhesus macaques found that these monkeys could use probability information to infer the most likely outcome of a random lottery, demonstrating an ability to predict future events based on observed statistical properties (De Petrillo & Rosati, 2019). In humans, the perception of entropy plays a role in decision-making within economic and financial contexts, as Chen (2011) discusses how an entropy theory of mind might influence market behavior and investor psychology. Furthermore, Wasserman et al. (2004) discuss variability discrimination in humans and animals, emphasizing the universal importance of detecting and interpreting variability and randomness across different contexts. This capability for entropy detection can assist in the efficient allocation of cognitive and physical resources, influencing survival and prosperity.

Entropy is a fundamental concept originally rooted in thermodynamics and statistical mechanics, but it has broad implications across various scientific disciplines, including information theory and psychology. In thermodynamics, entropy is defined as the amount of energy in a system that is not available for doing useful work, often manifesting as the system's tendency to move towards chaos or equilibrium (Clausius, 1862). This concept was extended to information theory by Claude Shannon, who defined entropy as a measure of uncertainty or unpredictability in the information content of a message (Shannon, 1948). Shannon's entropy is quantified by the formula

$$H = -\sum_{x} p(x) \log p(x)$$

where p( ) is the probability of occurrence of the message . Here, entropy measures the average amount of information produced by a stochastic source of data, essentially quantifying how much information is produced on average for each message sent from the source. In more general terms, entropy is often used to describe the randomness or complexity within a system, whether it be the arrangement of molecules in a physical system or the unpredictability of bits in a data stream.

Our work falls under the general domain of asking people to make judgments about

statistical properties of events, a skill demonstrated by the ability to discern frequencies, proportions, and contingencies among observed events (Peterson & Beach, 1967). This ability extends to the realm of entropy, where sensitivity of the randomness or unpredictability of events parallels judgments about statistical variance in a distribution (Miller, 1956). In our experiment, participants, who were recruited from Amazon's Mechanical Turk, listened to eight-second note sequences and assessed whether the notes in the sequences were mostly the same or different. We manipulated the entropy of pitches by varying the frequency of occurrence according to a discrete probability distribution, which systematically transitioned from fully random (uniform distribution) to highly predictable (single very high probability mode). This manipulation aims to probe participants' sensitivity to changes in entropy. The second manipulation is whether or not providing participants with feedback regarding their judgments. After each trial, participants received immediate feedback indicating whether their answer was correct, allowing them to compare their perception against the reality. This feedback mechanism was hypothesized to enhance learning and adaptation over time. By incorporating feedback, we aim to investigate how this information affects participants' ability to detect and adapt to changes in entropy levels. Our psychophysical analysis of the data indicated that participants could indeed assess entropy, and we proposed a series of cognitive models to interpret this ability.

Experiment 1A

The purpose of Experiment 1a was to validate the general procedure and psychophysical approach for measuring participant sensitivity to variations in entropy between collections of note sequences. In our experiment, entropy represents the randomness of pitch events within a sequence of 64 notes. We calculated the minimal and maximal entropy levels for these sequences and selected ten intermediate entropy values at equal intervals, creating a total of 12 distinct entropy values to present a spectrum of randomness in the auditory stimuli. The procedure, modeled after Allan et al. (2008)'s Experiment 1c, involved fitting psychophysical curves to group data, rather than individual subject data, allowing us to assess general trends in perceptual sensitivity to entropy.

Methods

Participants

For experiment 1A, A total of 46 participants were recruited from Amazon's Mechanical Turk. Mean age was 50 (range = 34 to 70). There were 23 females, and 23 males. There were 42 right-handed participants. 23 participants reported normal vision, and 23 participants reported corrected-to-normal vision. 45 participants reported English as a first language, and 1 participant reported English as a second language. 24 participants were randomly assigned to the feedback condition, and 22 participants were randomly assigned to the no feedback condition.

Apparatus and Stimulus

The experiment was programmed in JavaScript, HTML and CSS using JsPsych (De Leeuw, 2015). Just Another Tool for Online Studies (JATOS) was used to host this experiment online (Lange et al., 2015). The source code for the experiments is available from the project website.

The stimuli consisted of 96 digital audio pieces, generated using a combination of R and Python packages. Each piece of audio comprised 64 note events across eight pitches within an octave, synthesized using the R libraries Midiblender and Pyramidi. These were played at a tempo of 16th notes at 120 BPM, or eight notes per second with each sequence lasting 8 seconds in total. The audio was then converted to mp3 files by a R library Fluidsynth, set to a piano voice. Note velocity was set to 120 for all notes, and participants were advised to tune the volume to their comfortable listening level. The audio files can be accessed via the project website.

The 96 audio pieces were categorized into 12 groups, each containing eight pieces with identical entropy level. The 12 entropy values were evenly spaced across a range from the minimum of .87 bit to the maximum of 3 bits. An algorithm was developed to randomly generate the note sequences for the audio. This algorithm started from the most unbalanced distribution in which a pseudo-pitch was repeated 57 times and the other seven pseudo-pitches appeared once each. It proceeded by dividing the pseudo-pitches into two categories: one with the highest occurrence and the other comprising the remaining pseudo-pitches. The algorithm then randomly

selected a pseudo-pitch from each group and exchanged one occurrence between them. This process of selection and exchange was repeated until reaching the distribution with the minimal entropy. Consequently, a list of partitions with varying entropy levels was created. From this list, 12 partitions were selected for their entropy values, which were closest to the 12 predefined evenly spaced intervals between the minimum and maximum entropy values. For each chosen partition, eight distinct note sequences were created, with each sequence representing an individual audio piece. This was achieved by randomly assigning pitches to the pseudo-pitches and shuffling them. The script of this procedure is accessible on the project's website.

Design and Procedure

The study used the method of constant stimuli method to explore entropy sensitivity. The primary manipulation was the amount of entropy in a musical sequence across 12 parametric steps, ranging from .87 to 3 bits. A secondary manipulation was whether or not participants received feedback on their judgments after each trial. Participants were required to access whether a sequence was mostly same vs mostly different after listening to it entirely. There were 8 judgments for each of 12 entropy levels, culminating in a total of 96 trials.

Participants discovered our experiment on the Amazon Mechanical Turk platform (Crump et al., 2013) by searching for Human Intelligence Tasks (HITS) which listed our study as an available task. By selecting the task, participants were directed to our website, starting with a welcome page. The page informed them that the task would take approximately 15-20 minutes and required a desktop or laptop computer with Chrome or Firefox browsers and an audio device. After clicking Continue on the welcome page, participants signed a consent form on the second page and completed a demographic questionnaire on the third page. This questionnaire collected data on age, sex, handedness, vision, and proficiency in English.

Following the questionnaire, participants received instructions explaining that they would listen to short audio sequences varying in the amount of repeated notes. Their task was to determine whether the sequence contained mostly the same notes or mostly different notes. The instructions included a directive to adjust the audio volume and provided two examples of note

sequences to clarify the task. One with mostly the same notes and another with mostly different notes. Each example was explicitly described to illustrate these properties. The instructions set the stage for the 96 trials, which were randomly ordered by the experiment's software.

Each trial began with a white empty screen for 500 milliseconds, followed by an 8-second audio and a brief notice about the task. The response options, "Mostly Same" or "Mostly Different," were enabled only after the audio playing ended. For participants in the feedback condition, they received a one-second message indicating "Correct" or "Incorrect" after each trial. This feedback was based on a comparison of their choice to a preset criterion, which defined sequences with more than 2 bits of entropy as "mostly different" and those with less as "mostly same." Participants in the no-feedback condition did not receive this message. The trials continued sequentially until completion.

After all trials were completed, a survey was administered, asking participants open-ended questions about any issues with the audio, the process, and their overall experience. The study concluded with a debriefing session for the participants.

Planned Analysis

We planned to categorize the collected data into two groups—feedback and no-feedback—and then use Quickpsy (Linares & López-Moliner, 2016) to fit psychometric functions to this grouped data, estimating the threshold and slope within a basic signal detection framework as outlined by Kingdom and Prins (2016). This framework assumes that the perception of stimuli adheres to a normal distribution and maintains consistent standard deviations across all conditions.

Results

For each feedback condition in E1A, all participants were treated together as a single participant. Group-level counts of judging "mostly different" for each level of bits were fit with standard normal curves using the R package 'quickpsy' (Linares & López-Moliner, 2016). Figure Figure 1 displays proportion "mostly different" as a function of entropy for each feedback condition, along with fitted curves.

Figure 1

Proportion "mostly different" as a function of entropy for each feedback condition, along with fitted curves.
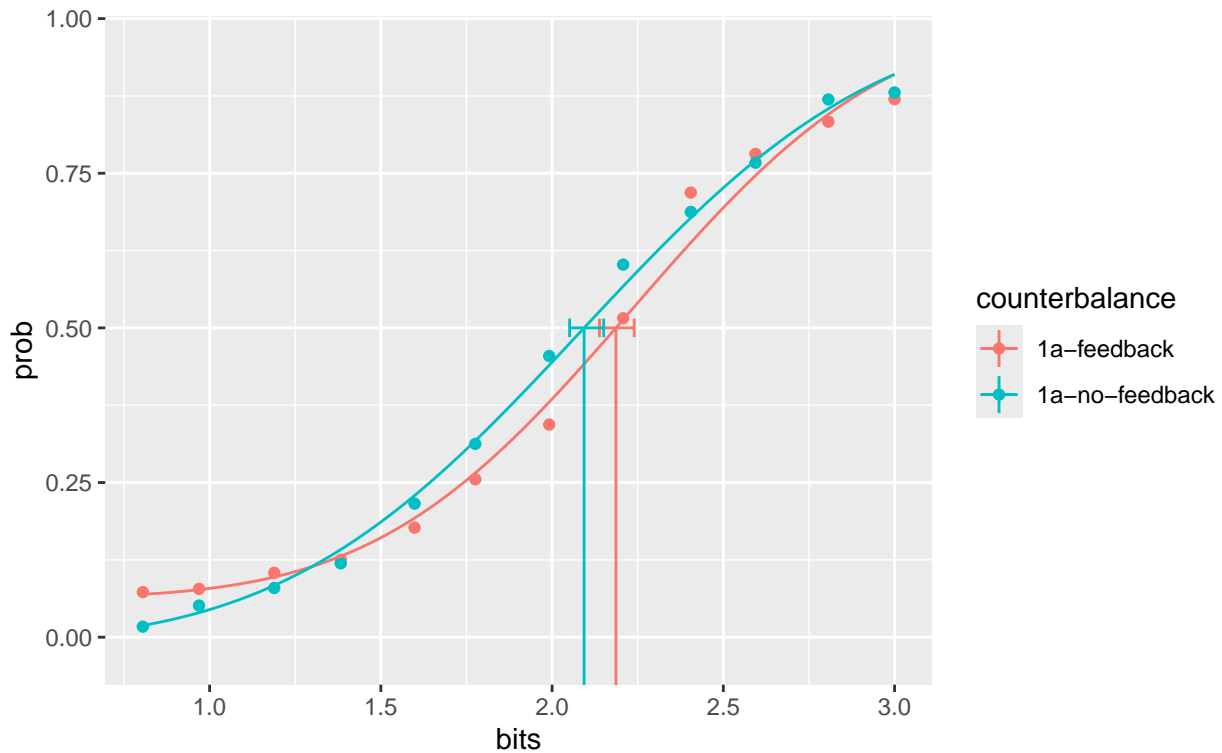
The point of subject equality (PSE) is the location on the curve where participants respond "mostly same" vs "mostly different" with equal probability (50%). The PSEs for each condition were $PSE_{feedback} = 2.19$, and $PSE_{nofeedback} = 2.09$.

General Discussion

Our study designed a psychophysical experiment to investigate participants' sensitivity to entropy or randomness in pitch sequences. The primary independent variable was the entropy of the pitch sequences, which ranged from .89 to 3 bits. The secondary independent variable concerned whether participants received immediate feedback on their performance after each trial. A psychometric function was fit on the data result, that illustrated that participants could accurately distinguish between levels of entropy among pitch sequences and placed the criterion around the midpoint of 1.95 bits. The criteria for discrimination were 2.19 bits with feedback and

2.09 bits without feedback. The slope of the psychometric function, indicating the sensitivity of perception, was 0.59 with feedback and 0.68 without feedback.

How did participants perceive the entropy and make their decisions? We explore a variety of potential interpretations that could explain this phenomenon. These interpretations may or may not provide a comprehensive understanding of the underlying cognitive processes involved.

Candidate Model 1 - Explicit Counting Model

Participants approach the task similarly to how a computer program might, by explicitly counting individual notes as a note sequence plays and then explicitly computing entropy with the mathematical formula. Although it is not impossible for someone to adopt this method, it is highly unlikely that most participants used this strategy for the following reasons. First, it is not clear that people are capable of accurately counting separate occurrences of 8 different pitch events played at the rates in our experiment. Second, it is not clear that our participants had any explicit knowledge of Shannon's entropy formula, and as a result they could not have computed entropy in such an explicit manner.

Candidate Model 2 - Prominent Pitch Model

Participants may simplify the task by focusing on the most prominent pitch during the audio playback, estimating its frequency, and then making their decision based on this estimate. After the sequence concludes, they compare the estimated frequency to a mental criterion to judge the level of entropy. This strategy relies on identifying and evaluating the most frequently occurring pitch, rather than attempting to count all pitches. Notably, one of the experimenters employed this strategy during a pilot test. He successfully identified the prominent pitch, estimated its frequency, and then made a judgment accordingly. This anecdotal evidence suggests that this strategy is not only plausible but also potentially effective.

To further validate this model, we developed a computer simulation using the Minerva Model within the framework of Instance Theory. The Minerva Model posits that memory stores individual instances of experiences which are accessed when similar instances are encountered (Jamieson et al., 2022). In this simulation, all pitches were assembled into a probe for testing

individual notes, allowing us to calculate echoes. These echoes were then used to determine the proportions of different pitches. The highest proportion was compared to a predefined criterion to assess if the audio exemplified low entropy. Remarkably, when noise was disabled from the simulation, it achieved 100% accuracy across all 96 stimuli.

If the model holds true, a specific prediction can be made about participants' judgments under certain conditions. When presented with a sequence where two pitches occur with high frequency, resulting in a relatively low overall entropy, participants using this strategy might mistakenly assess the sequence as having high entropy. This misjudgment would occur because the focus is on the prominence of individual pitches rather than the overall distribution of all pitches.

Candidate Model 3 - Continuous Assessment Model

This model, informed by the findings of Rebecca N. Faubion-Trejo (Faubion-Trejo & Mantell, 2022), posits that participants can effortlessly perceive both absolute and spectral content in "plinks," which are brief audio clips. During audio playback, listeners perceive a series of plinks and retain the information in memory. Each plink may prompt a momentary assessment of entropy, with certain plinks, such as those featuring repetitive notes or ending sequences, more likely to trigger this assessment.

The process of evaluating entropy also leverages the Minerva Model to interpret how memory influences perception. If a plink contains a dominant note, or if the last dominant note from a previous assessment is recalled, this note is used as a probe to test all remembered information related to the audio, resulting in a global activation value. A high activation value leads to a temporary judgment that the audio sequence is low entropy; a lower value suggests high entropy. After playback concludes, the most recent judgment solidifies into the final assessment of the audio's entropy.

This model mirrors a natural human tendency to make continuous, temporary judgments rather than a single, conclusive evaluation at the end. To validate this model, a computer simulation was implemented, achieving an accuracy rate of 99% across all experimental stimuli.

This encouraging result suggests potential validity of the Continuous Assessment Model in replicating human auditory judgment processes, although further studies are necessary to confirm the model.

An assumption of this model is that participants synthesize their judgments from individual perceptions of "plinks" rather than perceiving the audio sequence as a cohesive whole. If this assumption holds true, the specific positioning of prominent pitches within the sequence should significantly influence the overall judgment of entropy. To test this, we propose an experiment where we manipulate the locations of prominent pitches within audio clips that have a baseline entropy around the midpoint of 1.94. We will adjust prominent pitches to different positions, especially at the beginning and end of the sequences. This setup tests whether these positions significantly influence overall entropy judgments. If our model is correct, audio judged with unbalanced beginning and ending plinks should be perceived as more unbalanced. This would indicate that pitch locations critically affect entropy assessments, suggesting audio is perceived in segments. If pitch location manipulations have no impact, it would imply a holistic perception, challenging our model.

Candidate Model 4 - Cumulative Scoring Model

This model has a similar assumption as the previous model, that participants can assess the entropy of short audio segments, referred to as plinks. The key difference in this model is the method of aggregating these assessments. As each short audio segment is heard, it is scored based on its level of entropy and categorized as low entropy, middle entropy, or high entropy. Each segment's score adjusts a cumulative total score, which starts at middle entropy. This total is updated continuously as each new segment is assessed throughout the playback. At the conclusion of the audio, the final judgment regarding the overall entropy of the piece is based on this cumulative score. This model emphasizes a dynamic and ongoing process of evaluation, where the entropy judgment is the sum of sequential assessments, reflecting a granular approach to auditory perception. A key prediction of this model is that audio which is globally balanced but locally repetitive will be mistakenly judged as having low entropy. However, our daily experience

suggests that people do not always make this mistake.

Candidate Model 5 - Neural Simulation Amendment:

This model incorporates an artificial neural network (ANN) to simulate the calculation of entropy. Built using TensorFlow, this ANN model processes the frequency vector of pitches from audio data through a sequence of densely connected layers. It utilizes ReLU activations for the hidden layers and a linear activation for the output layer to predict the entropy value. The use of ANNs to simulate cognitive processes remains a contentious method; critics argue that while an ANN may mimic the output of human cognitive processing, it does not necessarily mirror the intricate mechanisms of the human brain (Donaldson, 2008).

All code of the mentioned simulations is available from the project website.

Conclusion

Our study designed a psychophysical experiment to investigate participants' auditory sensitivity to entropy in pitch sequences. We fitted a psychometric function to the data, indicating that participants could distinguish between various levels of entropy, with the decision criterion placed around the midpoint. While we explored several computational models to shed light on the cognitive processes involved in assessing entropy, definitive conclusions about these cognitive mechanisms could not be reached. Future research might focus on further distinguishing these cognitive processes from those involved in estimating proportions and deepening our understanding of the human capacity to assess entropy. Additionally, an archival study comparing visual and auditory perceptions of entropy could offer valuable insights into how these processes may differ across sensory modalities.

References

Allan, L. G., Hannah, S. D., Crump, M. J. C., & Siegel, S. (2008). The psychophysics of contingency assessment. *Journal of Experimental Psychology: General*, *137*(2), 226–243. https://doi.org/10.1037/0096-3445.137.2.226

Chen, J. (2011). The entropy theory of mind and behavioral finance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1734526

Clausius, R. (1862). Ueber die Wärmeleitung gasförmiger Körper. *Annalen der Physik*, *191*(1), 1–56. https://doi.org/10.1002/andp.18621910102

Cristín, J., Méndez, V., & Campos, D. (2022). Informational entropy threshold as a physical mechanism for explaining tree-like decision making in humans. *Entropy*, *24*(12), 1819. https://doi.org/10.3390/e24121819

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS One*, *8*(3), e57410.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.

De Petrillo, F., & Rosati, A. G. (2019). Rhesus macaques use probabilities to predict future events. *Evolution and Human Behavior*, *40*(5), 436–446. https://doi.org/10.1016/j.evolhumbehav.2019.05.006

De Pinho, M., Mazza, M., Piqueira, J. R. C., & Roque, A. C. (2002). Shannon's entropy applied to the analysis of tonotopic reorganization in a computational model of classical conditioning. *Neurocomputing*, *44-46*, 359–364. https://doi.org/10.1016/S0925-2312(02)00382-X

Donaldson, S. (2008). A neural network for creative serial order cognitive behavior. *Minds and Machines*, *18*(1), 53–91. https://doi.org/10.1007/s11023-007-9085-z

Faubion-Trejo, R. N., & Mantell, J. T. (2022). The roles of absolute pitch and timbre in plink perception. *Music Perception*, *39*(3), 289–308. https://doi.org/10.1525/mp.2022.39.3.289

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology*, *1*(3), 174–183. https://doi.org/10.1038/s44159-022-00025-3

Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: A practical introduction*. Elsevier Science & Technology. http://ebookcentral.proquest.com/lib/brooklyn-ebooks/detail.action?docID=4332363

Lange, K., Kühn, S., & Filevich, E. (2015). " just another tool for online studies"(JATOS): An

easy solution for setup and management of web servers supporting online studies. *PloS One*, *10*(6), e0130834.

Linares, D., & López-Moliner, J. (2016). Quickpsy: An r package to fit psychometric functions for multiple groups. *The R Journal*, *8*(1), 122. https://doi.org/10.32614/RJ-2016-008

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81. https://psycnet.apa.org/record/1957-02914-001

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*(1), 29–46. https://doi.org/10.1037/h0024722

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks*, *18*(3), 225–230. https://doi.org/10.1016/j.neunet.2004.12.004

Wasserman, E. A., Young, M. E., & Cook, R. G. (2004). Variability discrimination in humans and animals: Implications for adaptive action. *American Psychologist*, *59*(9), 879. https://doi.org/10.1037/0003-066X.59.9.879