Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

Predictive coding of multisensory timing

Zhuanghua Shi [1] and David Burr[2,3]

The sense of time is foundational for perception and action, yet it frequently departs significantly from physical time. In the paper we review recent progress on temporal contextual effects, multisensory temporal integration, temporal recalibration, and related computational models. We suggest that subjective time arises from minimizing prediction errors and adaptive recalibration, which can be unified in the framework of predictive coding, a framework rooted in Helmholtz's 'perception as inference'.

Addresses
[1] Department of Psychology, University of Munich, Munich, Germany
[2] Neuroscience Institute, National Research Council, Pisa, Italy
[3] Department of Neuroscience, University of Florence, Florence, Italy

Corresponding author: Shi, Zhuanghua (strongway@psy.lmu.de)

Current Opinion in Behavioral Sciences 2016, **8**:200–206

This review comes from a themed issue on **Time in perception and action**

Edited by **Warren H Meck** and **Richard B Ivry**

For a complete overview see the Issue and the Editorial

Available online 17th February 2016

http://dx.doi.org/10.1016/j.cobeha.2016.02.014

Introduction

The sense of time, unlike other senses, is not generated by a specific sensory organ. Rather, all events that stimulate the brain, regardless of sensory modality, contain temporal cues. Because of heterogeneous processing of sensory events, subjective time may differ significantly for a given duration across modalities. For example, an auditory event is often perceived longer than a visual event of the same physical interval [1]. Subjective time is also susceptible to temporal context, voluntary actions, attention, arousal and emotional states, all of which can bias it away from physical time [2,3,4••,5,6]. Over the past several decades, researchers have advanced our understanding of how we perceive and integrate multisensory and sensorimotor timing, with examples such as the 'central-tendency' effect [7••,8••,9], the time shrinking illusion [10], and sensorimotor temporal recalibration [11,12••]. In this article we examine a few selected duration-related temporal phenomena and related computational models, and show how those phenomena can be parsimoniously explained within the predictive coding framework [13,14,16•]. We propose that subjective time is an outcome of adaptive processes of the brain that minimize the overall estimation error to boost the reliability of estimation of external temporal structures.

Subjective time as inference

One and a half centuries ago Hermann von Helmholtz famously suggested that perception can be understood as a process of *unconscious inference*: "The connection between the sensation and external object can never be expressed without anticipating it already in the designation of the sensation... This is because inductive reasoning is the result of an unconscious and involuntary activity of memory" [17]. Time perception is also the result of unconscious inference. Subjective time can be easily influenced by internal expectation, as suggested by Karl Vierordt [18] around the same time as von Helmholtz. He observed that subjective judgment of duration is attracted to an 'indifference point', which is close to the central mean of all the durations experienced [9,18]. That is, short durations tend to be overestimated and long durations underestimated. Hollingworth later coined this phenomenon of gravitation toward the expected mean magnitude as the '*central tendency*' *effect* [19].

The recent surge of interest in the central tendency effect [7••,8••,20•,21,22] has taken this topic to a new level within Bayesian inference framework. This development has been motivated by the fact that, across a wide variety of tasks, the fundamental problem encountered by the brain is coping with uncertainty [15]. To minimize uncertainty, the brain needs to maximally utilize the available information, combining not only sensory input but also top-down 'prior belief' in a weighted average manner. In Bayesian terms, perception emerges from probabilistic inference, including the likelihood associated with the sensory evidence and prior belief (see Box 1). While this type of weighted average is clearly beneficial when the external environment is relatively stable, combining multiple sources of information in the brain would engender perceptual and cognitive biases when the environment changes.

Jazayeri and Shadlen [7••] recently reinvestigated the central tendency effect in duration reproduction using a Bayesian approach, and confirmed that the fundamental principle of central tendency is a strategy to minimize the overall temporal reproduction errors by combining both sensory likelihood and prior knowledge (e.g., the statistical distribution) of the to-be-estimated duration. Their approach is illustrated in Figure 1. When asked to reproduce temporal intervals, people tend to underestimate long intervals and overestimate short intervals, always 'regressing toward the mean'. Importantly, the mean is set dynamically, for the specific range being tested in that

Box 1 Dynamic updating of priors in time perception

Various types of contextual calibration of time perception can be explained in the framework of Bayesian inference(See Figure I) [4••]. The central idea of Bayesian inference is that the brain uses all available temporal information to minimize the prediction error. One source of temporal information comes directly from sensory inputs, which depends on sensory measures and signal quality. For example, for a given duration D, the sensory measure is S. This cue can be expressed in Bayesian term as the likelihood function $P(S|D)$. Another cue that the brain often uses is internal expectation based on the prior knowledge. In Bayesian term it is the prior function $P(D)$. According to Bayes' rule, the probability of a duration being D, given the sensory measure S is the product of the prior probability and the likelihood, normalized by the probability of the sensory measures:

$$P(D|\}\mathrm{S}) = \frac{\mathrm{P(S|\}D)P(D)}}{\mathrm{P(S)}}$$

The probability distribution $P(S|D)$ is known as the posterior probability. When both the likelihood and prior are independent Gaussians, the optimal duration estimate can be predicted by

$$D = wD_s + (1-w)D_p$$

where D_s and D_p are the expected mean of the likelihood and prior, and the weight $\mathrm{w} = (1/\sigma_s^2)/(1/\sigma_s^2 + 1/\sigma_p^2)$ is proportional to its reliability, in which $1/\sigma_s^2$ and $1/\sigma_p^2$ are the reliability of the likelihood and prior. The variance of this optimal estimate is $(\sigma_s^2\sigma_p^2)/(\sigma_s^2 + \sigma_p^2)$, which is the minimum variance among all possible linear weighted combinations between the sensory estimate and the prior. When there are two conditional independent likelihoods (e.g., one from the auditory modality and another from the visual modality), and the prior is not the focus factor, the optimal estimate is very similar:

$$D = w_a D_a + (1-w_a) D_b$$

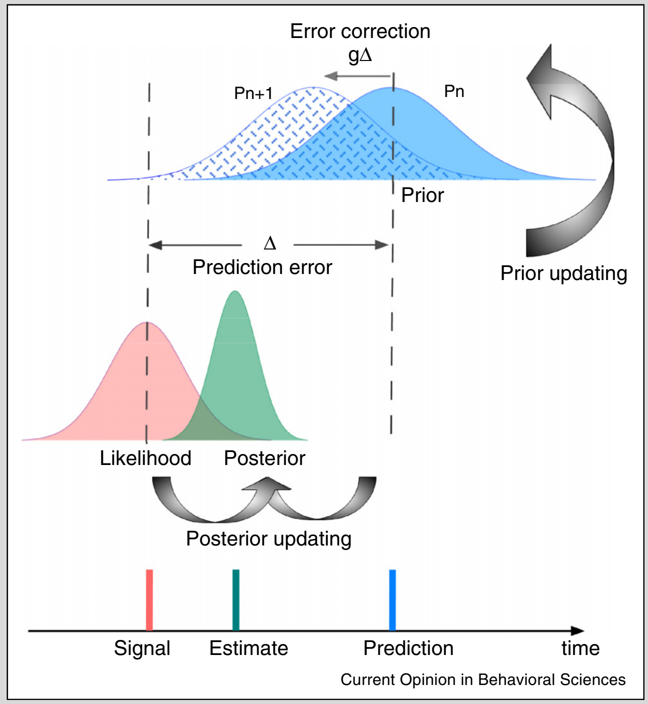
where D_a and D_b are the mean of two individual sensory estimates, and the weight w_a is proportional to its reliability.

With predictive coding, the internal predictive prior is not fixed, but is dynamically adjusted from the prediction errors. The top-down predictions are delivered through the backward connections. So long as this successfully predicts the lower level activity, all is well, and no further action needs to ensue. But where there is mismatch, a 'prediction error' occurs and the ensuing (error-indicating) activity is propagated to the higher level. This automatically adjusts probabilistic representations at the higher level so that top-down predictions cancel prediction errors at the lower level, yielding rapid perceptual inference [13–15,16•]. In a simple case, this predictive processing can be described by Kalman filter [4••,46] — a dynamic optimal prior updating process when noises are Gaussian:

$$P_n = (1-g)P_{n-1} + g\Delta$$

where P_n and P_{n-1} are the priors at time n and $n-1$, g is the Kalman gain, which is optimally determined by the variances of the internal prior and the prediction error. As shown by a developmental study on the temporal recalibration [12••], Kalman gain is larger in the adult group compared to the young groups (see text).

Figure I



Schematic illustration of Bayesian inference of duration. The red curve denotes the likelihood $P(S|D)$ for a given duration signal, the blue curve the prior at time n, and the dashed blue curve the updated prior at time $n+1$. The dark green curve is the posterior based on Bayesian inference. There are two updating processes: the posterior updating based on the cues and the prior is for reliable sensory estimates, and the prior updating based on error correction is for minimizing forthcoming prediction errors.

session. This is brought out most clearly for reproductions at 850 ms: the bias can be either toward shorter or longer intervals, depending on the range of intervals sampled in that particular session (lower panels).

Note that the internal prior may not equate to the rigid physical distribution of the stimuli, but rather be better captured as a smoothed approximation of the distributions up to third-order moments [8••,20•]. As sensory precision may vary among different groups of individuals as well as across different modalities [8••,21,23], central tendency effects vary according to the weighted average strategy of Bayesian inference. By testing subjects with various levels of musical expertise, Cicchini and colleagues [8••] have demonstrated the variation of tendency effects is closely related to Bayesian optimal encoding. Non-percussionists, who had large variability of visual duration reproduction, showed a standard central tendency effect, while expert drummers responded veridically owing to their high precision of reproduction (Figure 1). Similar variations of central tendency has been shown in patients with Parkinson's disease [23]. Patients are less prone to the central bias with their dopaminergic medication than without medication, as patients have higher sensory reliability in their medication state. Those findings [8••,21,23] suggest the brain represents recent statistics of event duration, and this information is incorporated in on-going perception, thus, producing biases such as regression to the mean; however, the degree of tendency biases depends crucially on the sensory reliability [8••].