Human Time Perception –
Predictable visual stimuli are perceived earlier than unpredictable events

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Barbara Wirxel
aus Warendorf, Deutschland

August - 2017

Tag der mündlichen Prüfung: 30.Oktober 2017

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: PD Dr. Axel Lindner

2. Berichterstatter: Prof. Dr. Christoph Braun

Prüfungskommission: PD Dr. Axel Lindner

Prof. Dr. Christoph Braun

Prof. Dr. Uwe Ilg

PD Dr. Gregor Hardieß

Erklärung / Declaration:

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

“Human Time Perception –
Predictable visual stimuli are perceived earlier than unpredictable events”

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, den ....
Datum / Date Unterschrift /Signatur

Table of contents

Abstract

What is it that enables us to timely react to visual events despite the significant processing delays within the visual system? This delay is estimated to be already about 100ms in higher visual areas, a delay that is relevant if one needs to initiate fast reactions, such as catching a ball in flight or initiating an escape. To compensate for delays in motor behavior, the brain employs predictive mechanisms. We aimed to investigate whether predictability of a visual stimulus not only affects behavior but also the time of perceived stimulus onset in humans. Specifically, we hypothesized that predictable visual stimuli have an earlier perceived onset than non-predictable stimuli do.

Our approach was the following: Subjects viewed streams of individual letters with a 1000ms standard interval between letters. This sequence of letters was either in alphabetical order, and thus predictable, or alternatively, the last letter of a sequence was chosen at random and thus not predictable. In each trial, subjects had to indicate whether or not the last letter agreed with the alphabetical order. Moreover, subjects had to estimate whether the duration of the last test interval, which was of varying length, was either longer or shorter than the standard interval. Varying the length of the last interval allowed us to estimate the point of subjective equivalence (PSE) between test and standard intervals. As we expected predictable letters to be perceived earlier, the PSE should be larger in predictable sequences than in non-predictable ones. In other words, the test interval would need to be relatively longer in order to compensate for the earlier perceived onset of predictable as compared to unpredictable letters.

Measurements of the PSEs confirmed our expectations and suggest that predictable visual stimuli are perceived earlier than non-predictable ones. Hence, even the perceptual system is compensating for delays in sensory information processing, allowing us to establish a timely perception of our environment.

To shed light on the related neuronal correlates of delay compensation, we performed a magnetoencephalography (MEG) study while subjects performed the same task. We intended to investigate relative temporal differences between predictive and non-predictive visual evoked potentials (VEPs) in our MEG study. Analysis of MEG data suggests that participants did generate predictions. This is indicated by a late signal difference between the non-predictable and the predictable

stimuli. However, we found no evidence for predictable stimuli evoking an earlier (or higher) peak of VEPs than unpredictable stimuli. Ultimately, it remains open whether the earlier perception of predictive vs. non-predictive stimuli is mediated through sensory prediction, through an interplay of prediction and postdictive perceptual evaluations, or through prediction and some yet unknown delay compensation mechanism.

Abbreviations

CDF	cumulative distribution function
CNS	central nervous system
EEG	electroencephalography
ERP	event-related potential
FEF	frontal eye field
FLE	flash-lag effect
fMRI	functional magnetic resonance imaging
fT	femto Tesla
Hz	Hertz
ICA	independent component analysis
IR	infrared / infrared-lab
ISI	inter stimulus interval
JND	just-noticeable difference
LGN	lateral geniculate nucleus
LGNd	dorsal lateral geniculate nucleus
M	magnocellular
MCS	method of constant stimuli
MEG	magnetoencephalography
MMN	mismatch negativity
ms	millisecond
MST	medial superior temporal area
MT	middle temporal area
OSR	omitted stimulus response
p	page
P	parvocellular
PEST	parameter estimation by sequential testing
PSE	point of subjective equivalence
RM	representational momentum
RMS	root mean squared
s	second
SQUID	superconducting quantum interference device
TOE	time order error

V1–V4	visual cortical areas 1–4
VEP	visual-evoked potential
vMMN	visual mismatch negativity
vs.	versus

Introduction

When a visual stimulus hits the retina, the resulting percept is significantly delayed due to the interleaved stages of neural processing. This delay is in the range of 80–100 ms for macaques (areas V3, MT, MST, FEF, V2, Schmolesky et al., 1998) (see Figure 1). We assume similar latencies for the human visual system.

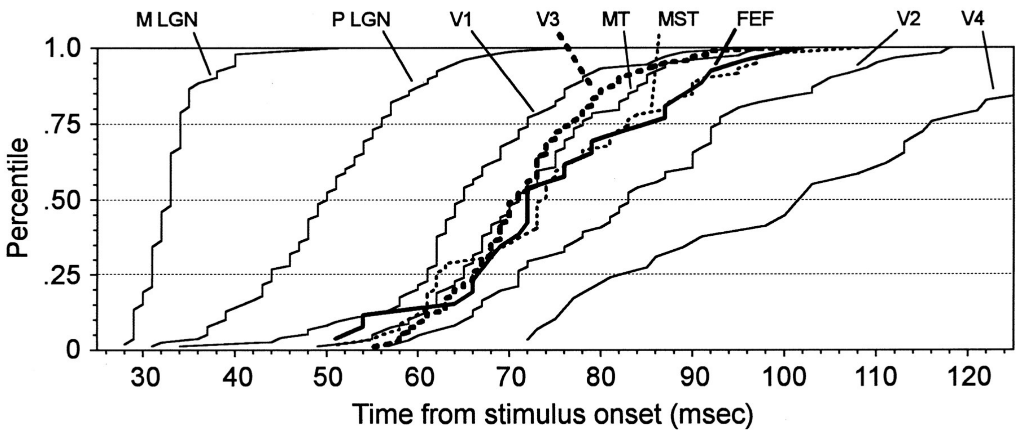

Figure 1: Timing of visual-evoked onset response latencies in Macaques.
"Cumulative distributions of visually evoked onset response latencies in the LGNd, striate and extrastriate visual areas as labeled. Percentile of cells that have begun to respond is plotted as a function of time from stimulus presentation. The V4 curve is truncated to increase resolution of the other curves; the V4 range extends to 159 ms" (Figure and caption taken from (Schmolesky et al., 1998).[1]

Although one-tenth of a second might not seem much at first glimpse, it is of ecological relevance. In order to interact with our environment, we need to be able to react fast to sensory information. For example, a fast-moving object such as a ball thrown in a game like cricket can move at a speed of up to 90 mph. The batsman has very little time to decide how to hit the ball and how to move his head out of the way if necessary (Nijhawan, 2008). Temporal accuracies better than 30 ms—sometimes even better than 5ms—are often required in such fast ballgames (Tresilian, 1993). Even when not playing high-speed ballgames, temporal accuracy[2] is relevant.

---

[1] Permission from The American Physiological Society is not required for use of figure 1 in thesis/dissertation according to permission request.

[2] Accuracy: Closeness of agreement of measurement results and the true value. This means that a measured value close to the true value is of high accuracy. "Accuracy" is also often termed as "bias" in statistical literature. It needs to be distinguished from the term "precision." Precision is the closeness of agreement between measured values obtained by replicate measurements. "Precision" is also often termed as "variability" in statistical literature.

Imagine a person walking at a speed of 3.6 km/hour, which equals 1 m/s. The person would misperceive the world by 10 cm due to a 100 ms neuronal processing delay. Any object perceived to be within 10 cm along the person's direction of motion would already be behind the person by the time she/he perceives it (Changizi, 2001). Both examples—interaction with fast-moving objects as well as slow locomotion—illustrate the ecological relevance of a neuronal processing delay in vision. This raises the question of whether and how neuronal processing delays are compensated.

Motor Prediction

One way to compensate for neuronal processing delays is motor prediction/motor anticipation. The word "prediction" is used in terms of estimating the future state of a system; such prediction may be realized by so-called "internal models." Internal models are neural representations that simulate motor-to-sensory transformations and vice versa (Ito, 2008; Wolpert & Ghahramani, 2000). For example, a copy of the arm's motor command (efference[3] copy[4]) is used in an internal forward model to simulate the interaction of the motor system (e.g. the arm) and the world, and can therefore predict behaviors/sensory feedback (see Figure 2). The opposite transformation—from desired sensory state to motor action—is simulated by the CNS as well, and is termed the inverse internal model. Inverse models are used to identify the respective control policy (motor command) needed to realize a desired action effect. Delay compensation might thereby be an integral part of inverse internal models. Similarly, an internal forward model can help the CNS perform movements precisely, without relying on (slow) sensory feedback from the limbs (Ito, 2008; Wolpert & Ghahramani, 2000).

If the parameters of both models are correct, the output of the forward model (predicted behavior) equals the input to the inverse model (desired behavior) (see Figure 2). Further on, we discuss forward models in the context of motor prediction.

---

[3] Efference: this term refers to outflowing neural signals from the central nervous system to the periphery, e.g. to central motor commands addressing peripheral muscles.

[4] Efference copy (also referred to as corollary discharge) is a copy of an efferent motor command flowing from the CNS to the muscles. The idea of an efference copy or corollary discharge was established in parallel by von Holst and Mittelstaedt (1950) and by Sperry (1950), respectively.

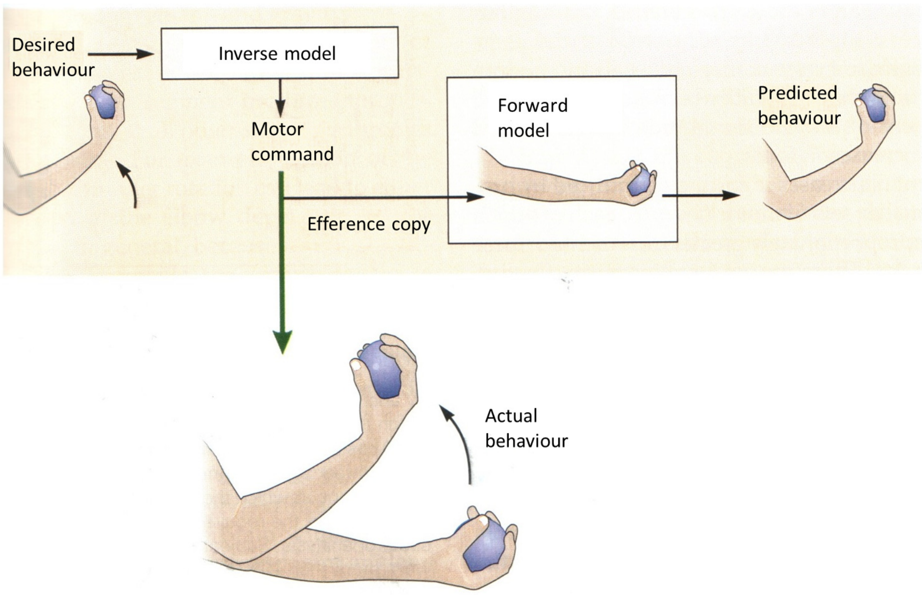

Figure 2: "Internal models represent relationships of the body and external world. The inverse model determines the motor commands that produce a behavioral goal, such as raising the arm while holding a ball. A descending motor command acts on the musculoskeletal system to produce the movement. A copy of the motor command—termed efference copy—is passed to a forward model that simulates the interaction of the motor system and the world and can therefore predict behaviors. If both forward and inverse models are accurate, the output of the forward model (predicted behavior) is the same as the input to the inverse model (desired behavior)" Figure and caption taken from Kandel et al., 2011[5].

The concept of motor prediction has already been postulated by Hermann von Helmholtz, who recognizes the need for the compensation of processing delays and suggests processes similar to those realized by forward control/forward models (Nijhawan, 2008). The effect of forward control can be demonstrated by our visual perception during eye movements; when we voluntarily move our eyes, the visual objects change retinal position, but we do not perceive the world as moving (Wolpert & Flanagan, 2001). It is suggested that this is because the "efference copy" of the active eye movement and the re-afferent retinal image motion exactly cancel out each other (von Holst and Mittelstaedt cited by Jeannerod, Kennedy, and Magnin, 1979, original is in German). In turn, when the eye is passively moved, without the use of eye muscles, we perceive an illusory displacement of the environment due to the missing efference copy, as the (passive) eye movement-induced retinal image motion is no longer perceptually cancelled. This situation can be easily mimicked in a self-experiment: When we cover one eye and touch the other eye's lid with a finger,

[5] With permission from McGraw-Hill Education.

the eyeball moves. Visual objects change position on the retina due to this movement. However, efference copy-based predictions about this visual motion are lacking as there is no motor command from the CNS to the eye muscles in this situation. This leads to a misperception of the world as moving (Wolpert & Flanagan, 2001). This experiment has already been reported by Helmholtz.

Motor prediction can be used in several ways. (i) As pointed out already, it can be used to approximate the sensory effect of movements and cancel this effect (based on efference copy and re-afference[6]). This is likely the case for voluntary eye movements. One possible reason such a predictive mechanism is implemented is the fact that sensory feedback-based compensation mechanisms are too slow, due to the neural processing delays involved.

(ii) Motor prediction is in fact used to maintain accurate performance despite feedback delays. Such a feedback delay appears between a motor command and its sensory consequences. Motor predictions compensating for such delays are omnipresent in everyday situations such as emptying a ketchup bottle (Wolpert & Flanagan, 2001). A person holds a ketchup bottle in one hand and tries to empty it by hitting the bottle's bottom with the other hand (see Figure 3). When holding the bottle with one hand and hitting it with the other, an appropriate grip force is applied. Grip force parallels load force with no delay. This is because an efference copy is used to predict the upcoming load force. In contrast, when another person hits the bottle, the force is externally applied (Figure 3, lower part).

(iii) Motor prediction can be used for motor learning. Discrepancies between actual and predicted outcomes can be used to update internal models, which would result in learning (Blakemore & Sirigu, 2003; Wolpert & Ghahramani, 2000).

Motor predictions are well-established and there is comprehensive literature on this topic—in particular on the role of internal models. Motor predictions are suitable for contributing to delay compensation, as exemplified in this section. However, they might not be the only elements that allow compensation of neural processing delays. Another kind of prediction might also contribute to this.

---

6 Afference is a neuron or neuronal pathway for incoming signals from the periphery to the central nervous system. Afferent signals can be distinguished according to the origin of sensory information—either from within the body (proprioceptive signals) or from the external environmental (exteroceptive sensory inputs). Both types of afferent information can be either self-generated (termed re-afference) or externally generated (termed ex-afference).

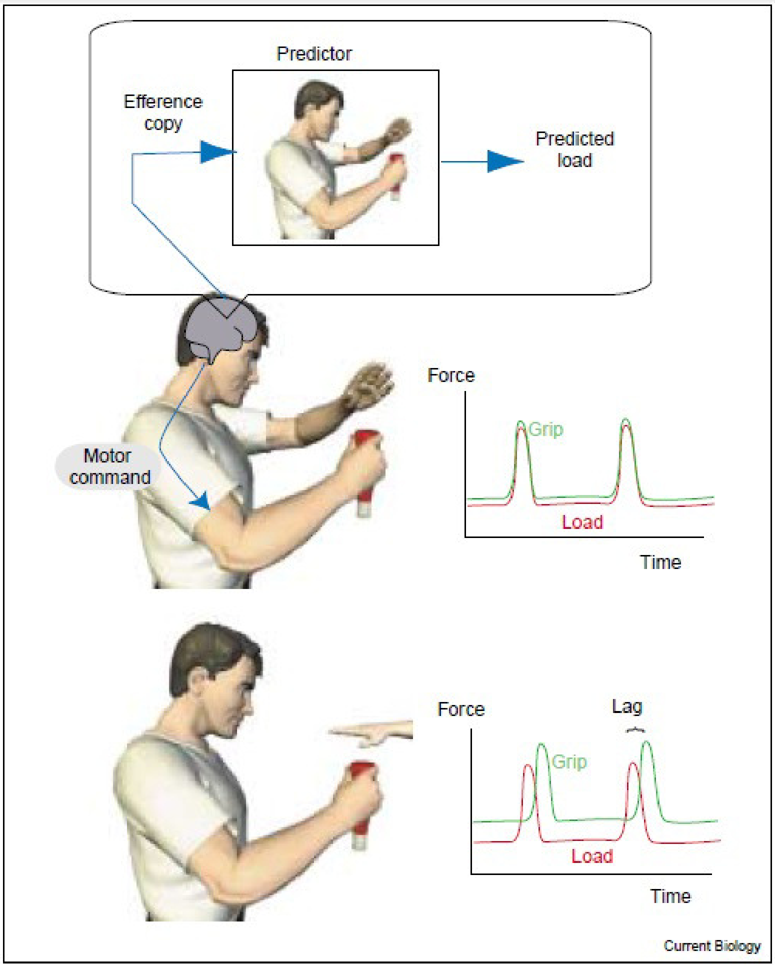

Figure 3: Schematic illustration of motor prediction and its contribution to compensating for feedback delays.
"To prevent a ketchup bottle from slipping, sufficient grip force must be exerted to counteract the load. When the load is increased in a self-generated manner (left hand strikes the ketchup bottle, top), a predictor can use an efference copy of the motor command to anticipate the upcoming load force and thereby generate grip which parallels load force with no delay. However, when the load is externally generated (another person strikes the bottle, bottom), then it cannot be accurately predicted. As a consequence, the grip force lags behind the load force and the baseline grip force is increased to compensate and prevent slippage." Figure and caption are taken from Wolpert and Flanagan, 2001[7].

Sensory Prediction

Another way to implement compensation for neuronal processing delays—apart from motor prediction—is sensory prediction. Prediction, as mentioned before, means to estimate the future state of a system.

[7] With permission from Elsevier, license number: 4153130736745.

Literature on sensory prediction is not as comprehensive as that on motor prediction. The idea that sensory processes may be per se predictive is not widespread and is under debate. Nevertheless, there are various groups that support the notion of sensory predictions (Cavanagh, 1997; Nijhawan, 1994, 2002).

The position that sensory processes are predictive is worth considering. It seems unlikely that sensory areas are excluded from prediction, because prediction is known to play a role in many different areas of the nervous system. Internal models, which serve as a predictive function in the motor area, might be generalized for the area of perception (Kawato, 1999; Nijhawan, 2008).

Large neural delays have been measured within the optic nerve and retina (Dreher, Fukada, & Rodieck, 2006; Ratliff & Hartline, 1959). Because vision is slow, a significant part of the delay during visual-motor interactions is caused by the visual process per se. Delay compensation within the visual system itself could thus provide a highly relevant contribution (De Valois & De Valois, 1991). Therefore, we take a brief look at evidence for delay compensation within the visual system, at the retinal level.

Sensory prediction at the retinal level: At the retinal level, temporal expectations about the visual world already exist. Mouse and salamander retinae detect and predict periodic patterns (Schwartz, Harris, Shrom, & Berry, 2007). In this study, an omitted stimulus response (OSR) is measured in retinal ganglion cell spike trains, revealing, among other things, the following: OSR is predictive for the time at which the next flash should have occurred instead of simply signaling that the flash sequence has ended. This is shown by OSR latencies, which are nearly constant for all presented stimulus frequencies (6–20 Hz). Mouse and salamander retinae are capable of recognizing and predicting periodic patterns. A violation of temporal expectation is already represented in the neural code at the retinal level.

Flash-lag Effect

The flash-lag effect (FLE) is an experimental paradigm that combines a discrete stimulus (a light flash) and a continuous stimulus (a moving object). The flash-lag illusion is often interpreted in the context of sensory prediction within the visual system. The FLE is designed to investigate whether visual responses to discontinuous stimuli are delayed, whereas visual responses to continuous (and thus predictable) stimuli are not. If this is the case, a paradigm concatenating continuous and discrete visual stimuli within the same display should reveal an anomaly. Indeed, the flashed object is perceived in a position lagging behind the perceived position of a continuous moving object, despite the two objects being spatially and temporally aligned. Numerous experiment designs have revealed the FLE. An illustration of such a design is given in Figure 4.

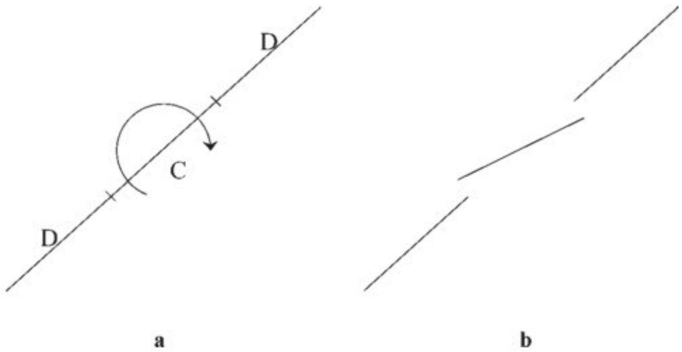

Figure 4: Schematic illustration of a possible layout of flash-lag paradigm.
"(a) In a dark room, a single physical rod, made of three segments, rotates clockwise in the direction of the arrow. The segment labeled "C" is illuminated with a continuous light source, while the discrete segments labeled "D" are illuminated with a brief flash produced by a stroboscope. (b) The percept of the observers." Figure and caption taken from Nijhawan 2008[8].

Three flash-lag conditions (cycles) are created to investigate the illusion in further detail (see Figure 5). These conditions are termed as standard flash-lag, flash-initiated, and flash-terminated conditions. In the standard flash-lag condition (complete cycle), a constantly moving object is visible before and after the flash. In the flash-initiated condition (initiated half-cycle), the flash occurs at the same time as the unpredictable onset of the moving object. The perceived FLE is comparable in magnitude to the standard FLE. In the flash-terminated condition (terminated half-cycle), a constantly moving object disappears unpredictably at the same time as the flash. Participants do not perceive the FLE in this condition. These results are surprising. One could expect the flash-lag illusion to be absent in the initiated condition and present in terminated condition; but this is not true in either case. Several different explanations for the FLE have been proposed. These include, among others, motion extrapolation, a predictive model, and a postdictive model. However, none of these approaches can by itself fully explain the FLE, including its presence in initiated conditions and its absence in terminated conditions.

[8] With permission from Cambridge University Press, license number 4160261278203.

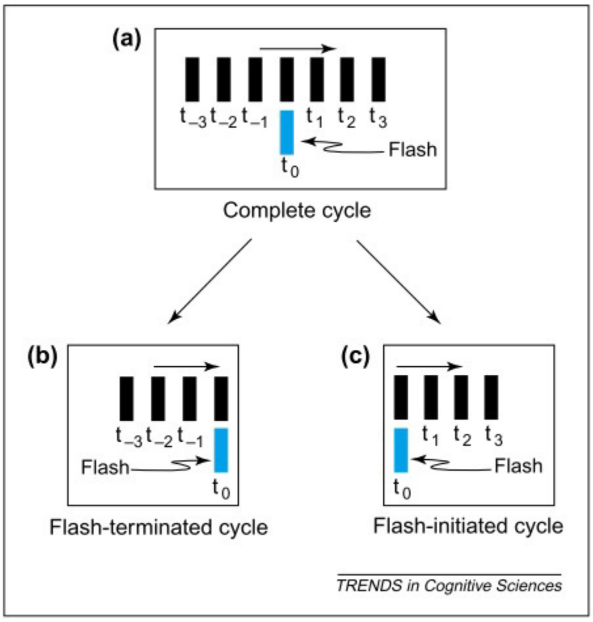

Figure 5: Schematic illustration of the flash-lag, split into flash-initiated, complete cycle, and flash-terminated.

"Half-cycle displays. (a) The standard (complete cycle) display with a moving bar (top) and a flashed bar (bottom) that had been used in flash-lag studies until the early 1990s. This display can be divided into two half-cycles. (b) In the flash-terminated half-cycle, motion offset occurs simultaneously with the flash. Observers perceived the moving object to stop in alignment with the flash (as it actually does). Thus, no flash-lag is observed. (c) In the flash-initiated half-cycle, motion onset occurs simultaneously with the flash. In this case, the flash-lag is perceived, and the effect is comparable to the complete cycle display […]." Figure and caption are taken from Nijhawan, 2002[9].

For further details on the variety of approaches and their contradictions, see (Eagleman & Sejnowski, 2000; Nijhawan, 2002, 2008). Two models are of particular interest in the context of this work—the predictive and the postdictive model. Therefore, I discuss in further detail these mechanistic interpretations of the FLE, while also highlighting supposed contradictions between these interpretations and empirical findings.

Flash-lag effect - predictive model: The motion extrapolation approach, as proposed by Nijhawan, is a predictive approach (Nijhawan, 1994, 2008). By extrapolating the trajectory of the moving object into the future, the predictive model compensates for neural delays. A retinal signal arising from the flash arrives after some temporal delay in the cortex. An object moves a distance of Δs in time Δt. The

[9] With permission from Elsevier, license number 4153180332238.

expected lag (∆s) between the perceived position and the physical position of the *moving* object is shortened by compensation, resulting in a corrected cortical spatiotemporal representation of the object. This corrected cortical representation of the *moving* object is ahead of the (uncorrected) cortical representation of the flash. The cortical representation of the flash cannot be corrected due to missing predictive information. This combination of one corrected and one un-corrected cortical representation elicits the flash-lag illusion (standard flash-lag).

The motion extrapolation model is not without contradiction, as it is incompatible with results revealed in the terminated condition. Compensation of the moving object should lead to a percept of object overshooting the termination point because the motion-termination is registered belated on cortical level. In fact, there is no percept of overshoot. The motion extrapolation model cannot explain the lack of FLE in the terminated condition and is therefore not suitable for fully explaining the flash-lag illusion (Nijhawan, 2002, 2008). In addition, the motion extrapolation model also does not explain the presence of the FLE for the flash initiated cycle. Flash and movement onset appear together at the cycle's initiation and are not predictable.

Flash-lag effect - postdictive model: Postdiction is the retrospective attribution of perceptions after the brain has collected information in a time window around the event (Nijhawan, 2008). The postdictive model in context of the FLE suggests that the flash interacts with current motion processing. In more detail, it assumes that the percept attributed to the time of an event—e.g. the flash—is a function of what happens within the ~80 ms time window following the event. Eagleman and Sejnowski (2000) designed an experiment in which pre-flash trajectories had no effect on the perceived position of a moving object, whereas different post-flash trajectories evoked different perceived positions of a moving object. Therefore, they argue that the FLE is not an extrapolation into the future, as suggested by the motion extrapolation and the predictive model; instead, it is a reconstructive interpolation of the past. To test how much/long information after the flash is integrated, they implemented a change in movement direction after the flash. Movement changes "up to 80 ms after the flash influenced the percept" (Eagleman & Sejnowski, 2000). Therefore, they assume a postdictive model to explain the FLE.

The postdictive model explains the absence of the FLE in terminated condition. This model predicts no FLE for the terminated condition, because the moving object no longer changes its position after the flash. However, the postdictive model has difficulties in explaining the standard FLE. A flash presented at t_0 is delayed for Δt (neural delay). The postdictive model assumes that the neural delay Δt is equally long for both flash and motion.

There is controversy as to whether the FLE provides evidence of a predictive process in vision, as well as whether it provides evidence of a postdictive process (e.g. Eagleman commented on Nijhawan's critique "Prediction and postdiction: Two frameworks with the goal of delay compensation" published as part of the "open peer commentary" section in Nijhawan 2008).

It should be noted that both views have the same goal: they try to place the perceptual location of a moving object closer to its physical location despite neural delays. To our knowledge, no model successfully explains the entire flash-lag illusion, standard, initiated, and terminated conditions.

Present Study

The aforementioned literature suggests that—apart from motor prediction—purely perceptual predictions could also contribute to delay compensation. However, to our knowledge, there is no compelling evidence for delay compensation at a strictly perceptual level.

Hence, for the visual domain, the question is whether predictability of a visual stimulus affects the time of perceived stimulus onset. Specifically, we hypothesize that predictable visual stimuli have an earlier perceived onset compared to non-predictable stimuli. First, we performed a series of psychophysical experiments that allowed us to develop an adequate design to probe our hypothesis at a behavioral level. Second, we collected magnetoencephalography (MEG) data while subjects performed the same task.

By exhibiting the underlying neuronal correlates of this delay compensation, we aimed to shed light on the underlying neuronal mechanism. It is possible that prediction and postdiction do not exclude each other. Instead, prediction and postdiction could interact. Supposedly, prediction can have an effect on postdictive processes. The existence of a predictive process is mandatory—e.g. during flash-lag illusion (standard flash-lag). Prediction can have a direct effect on perception. It could possibly interact with postdiction and have a joint effect on perception. For the present study, there was no necessity to make a commitment in regard to whether prediction has a direct impact on perception or a joint impact when combined with postdiction. Both approaches—postdictive and predictive—aim to place the perceptual onset of the last character closer to its physical onset. MEG data were analyzed by a master's student during lab rotation (Katrina Quinn, 01.09.2014–07.11.2014). The present thesis includes these MEG data.

Material and Methods

We tested our hypothesis in a visual psychophysical experiment. To come up with an adequate experiment design, several parameters were adjusted. Hence, we developed different designs, including an MEG-compatible version of the most promising design. The basic concept of all designs is described, along with the hypothetical results. Required adjustments are discussed in later sections.

Basic Concept/Paradigm

Figure 6 shows a schematic of the basic paradigm/concept of the experiment task. It is designed as a temporal judgment task (Wirxel & Lindner, 2012).

Participants were presented a string of individual characters in the middle of a black screen. Each character was presented for an equally long period ("presentation interval"). Characters were separated by inter-stimulus intervals (ISIs), namely standard intervals (or "fixed interval") and one "comparison interval." The comparison interval, which was always the last ISI, was of varying length/duration. This comparison interval is also referred to as "test interval." The sequence of characters was either in alphabetical order and thus *predictable* (Figure 6, upper part) or, alternatively, the last character of a sequence was chosen at random and thus *not predictable* (Figure 6, lower part). Note that all other characters (first to penultimate) followed the alphabetical order in both conditions. Character sequences varied from four to eight characters and sequences were allowed to start at any position in the alphabet that would allow a sequence to end at or before the letter "z." We expected participants to perceive the last character subjectively earlier in predictive condition than in non-predictive condition.

In each trial, subjects had to estimate whether the duration of the test interval—which was of varying length—was longer or shorter compared to the preceding standard intervals of fixed duration. Varying the length of the test interval allowed the estimation of the point of subjective equivalence (PSE) between test and standard intervals. During the response epoch, the subject had to first make a temporal judgment. We expected the difference in perceived onsets for predictive and non-predictive conditions to lead to different judgements of the temporal judgement task.

After responding to the temporal judgement task, subjects had to indicate whether or not the last character agreed with the alphabetic order (Wirxel & Lindner, 2012). Both tasks were answered by pressing one of two buttons of a response box. Subjects first had to perform a temporal judgement task by estimating that the test interval was either "too long" (right button press) or "too short" (left button press). Then they had to perform the control task (alphabetical order task). Subjects had to choose between "*Alphabet falsch beantwortet*" (alphabet answered incorrectly) or "*Alphabet richtig beantwortet*" (alphabet answered correctly). The option of "undecided" was not provided for either the temporal judgement or the control task. Subjects had five seconds to complete each of the two tasks (max. 10 seconds in total).

Predictive and non-predictive trials were randomly interleaved. A sequence length of four to eight characters was also chosen at random in order to prevent the subject from estimating the last character based on the first character of a sequence. With a fixed number of characters (fixed sequence length), this was made possible.

The control task (alphabetical order task) served to encourage subjects to focus on the order of the characters. Since the experiment was designed to address the difference between a predictable and a non-predictable condition, ignoring the alphabetic order could spoil the experiment. If a subject focused purely on rhythm and not on alphabetic order, we would not expect any difference between predictive and non-predictive conditions. To reveal whether subjects use predictive information, we made them focus on content by asking them answer a control question: "Was the last character correct according to the alphabetical order?" Only trials with correct answers to this question are taken into account for analysis.

It should be noted that the task design is not a reaction time task. Subjects are neither forced nor encouraged to answer as fast as possible. This is important, as the task was designed to address the *perceptual* level of the compensation mechanism for neuronal processing delay in the visual system, as explained in the introduction. We did not aim to investigate compensation mechanisms on sensorimotor level.

Hence, the task design does not require a reaction time task. Reaction times would possibly reflect sensorimotor learning.

Further, the last character for non-predictive condition was chosen at random. If we had always used a fixed character for unpredictable condition instead (e.g. “Z”), this last character would have been “predictable” in another way as well. This kind of predictability would be different compared to the predictability of a consecutive string of characters in alphabetical order. It would be predictable in shape—e.g. the shape of “Z”—if “Z” was always the last unpredictable character. Therefore, the last character for the non-predictive condition was chosen at random. Excluded options for this random choice were (i) the correct character according to the alphabetical order and (ii) the previous character. For example, for a sequence of four characters, the first of which were A, B, and C, the last character could be anything but “C” and “D.” As stimulus color, we chose red on a black background. Red has the fasted decay-rate on a CRT screen. In other words, the phosphorescence is the shortest with red color.

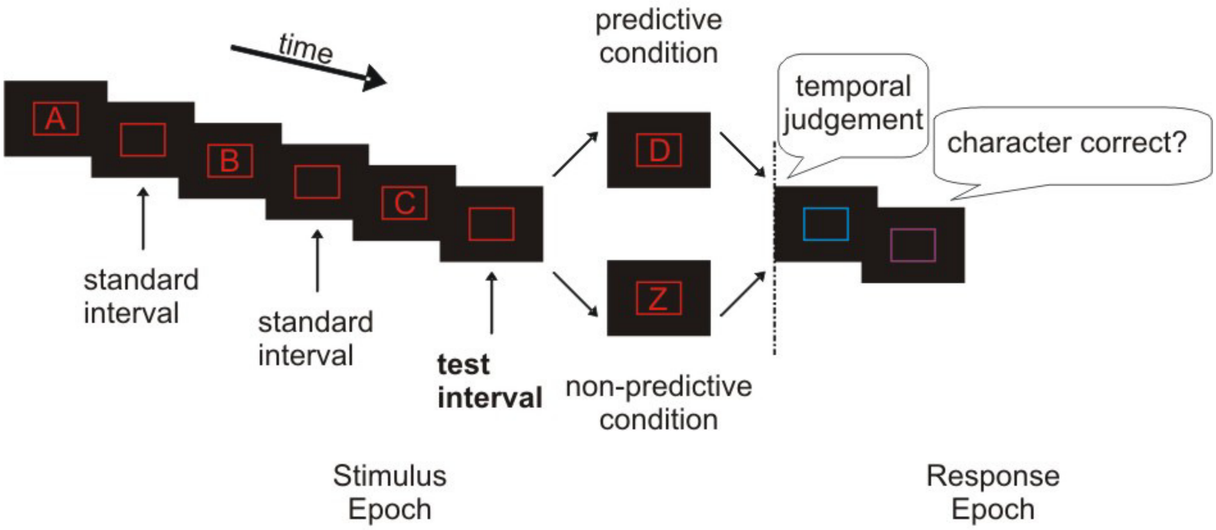

Figure 6: Experiment task design (basic concept).

Each character was presented for an equally long duration with a fixed standard interval in between characters. The sequence of characters was either in alphabetical order and thus predictable (character “D” in the figure) or, alternatively, the last character of a sequence was chosen at random and thus not predictable (character “Z” in this figure). The subjects’ first task was to judge the length of the test interval compared to the previous standard intervals (temporal judgement, indicated by the blue rectangle). The subjects’ second task was to indicate whether the last character was “correct” or “incorrect” according to alphabetical order (indicated by the purple/magenta rectangle).

The psychometric function is captured by a probit function, which is a cumulative density function (CDF) of a normal distribution. PSE is taken from the sigmoidal psychometric function. In our study, PSE refers to the point at which a comparison stimulus (varying last interval, “test interval”), is judged equal to the standard stimulus (first to penultimate interval of fixed length) on average. PSE cannot be measured directly and is given in milliseconds. It gives the length of the varying comparison interval at which it is perceived as “too long” in 50% of the cases and as “too short” in the other 50%, as compared to the standard interval. This is the interval length when a subject is undecided and is forced to judge either “too long” or “too short” by our experiment design.

As we expect predictable characters to be perceived earlier, the PSE should be larger in predictable sequences than the PSE in non-predictable ones. This is equivalent to the idea that at PSE, more time has to pass in the predictive condition as compared to the non-predictive condition in the variable delay.

Hypothetical Results

We expected subjects to perceive predictable characters earlier. Therefore, the PSE should be larger in *predictable* conditions than in *non-predictable* ones. This is equivalent to the idea that more time has to pass in the predictive condition as compared to the non-predictive condition. Therefore, the psychometric function should be shifted to the right in predictable stimuli (Wirxel & Lindner, 2012). In detail, we expect to reveal data as follows.

Figure 7 shows schematic of hypothetical results. The upper part shows the predictive condition, the middle part shows the non-predictive condition, and the lower part shows a combined graph illustrating the expected psychometric functions of both conditions. On the left, single trials for both conditions are displayed. On the right, temporal judgments of multiple trials are shown. For the predictive condition (Figure 7, upper left), a predictable last character (P) is displayed on the objective time – axis. A participant should perceive this character subjectively earlier, as shown on the subjective time – axis. Therefore, on a subjective time – axis, the predictable last character is located a bit on the left (i.e. at an earlier point in time).

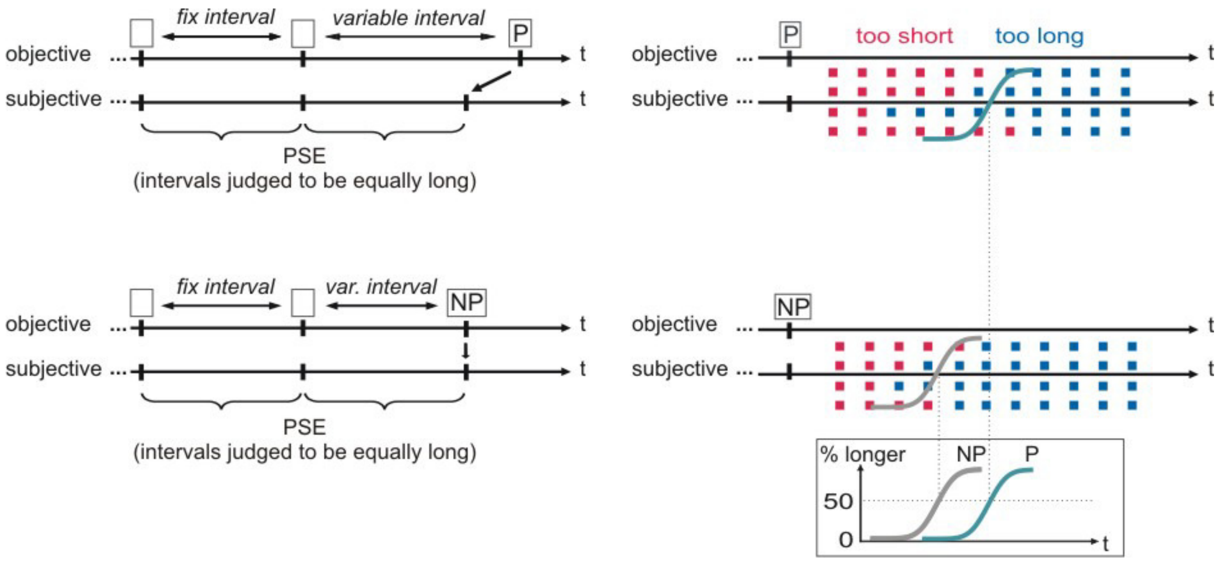

Figure 7: Schematic of expected results.

Upper part: Predictive condition, single trial on the left. Time between last predictable character (P) and penultimate character appears subjectively shorter than what the interval objectively was, as indicated by an arrow pointing to the left. Based on this shortened subjective time percept, subjects make their temporal judgment. Temporal judgments of multiple trials are displayed on the right. Each colored square represents one temporal judgment (red = too short, blue = too long). The interval length (objective time) at which a subject repeatedly judges 50% of the cases as too long (blue square) and 50% as too short (red square) is the PSE. PSE is visualized by the inflection of the sigmoidal psychometric function (greenish color for predictive case).

Middle part: Non-predictive condition. Single trial is shown on the left. Subjective time between last non-predictable letter (NP) and penultimate letter is as long objective time. Time percept is accurate, as indicated by an arrow pointing straight down. Based on this accurate subjective time percept, subjects make their temporal judgments. Temporal judgments are displayed on the right again. Square colors are defined as described above. PSE is visualized by the inflection of the sigmoidal psychometric function (greyish for the non-predictive case).

Lower part: Both psychometric functions (P and NP) are plotted in a combined graph. The predictive (P, greenish) function is shifted to the right. As mentioned before, the time between last predictable letter (P) and penultimate letter appears subjectively shorter than the objective interval (arrow pointing to the left). Therefore, more (objective) time has to pass in the P condition than in the NP condition to make subjects reach the PSE. On single-trial basis, this is indicated by the vertical misalignment of the last characters (P and NP) on objective time scales for both conditions (left upper and left middle part). On multiple-trial basis, it is indicated by the misalignment of the two psychometric functions (lower right). The predictive function is shifted to the right (larger PSE values) compared to the non-predictive function.

For the non-predictive condition (Figure 7, lower left), a non-predictable last character (NP) is displayed on the objective time – axis. The participant should perceive this last character accurately with respect to the time scale. Therefore, the non-predictable last character displays at the same location on both objective and subjective time axes.

For both conditions, we measured the PSE between standard (fixed) and test (variable) interval duration. Since fixed standard intervals were of the same length for both conditions, PSEs of the two conditions are comparable. PSE values give the objective, physical amount of time, which relates to the 50%-answer behavior.

For the predictive condition, more time has to pass during the variable interval to be judged as equally long as the standard interval. For the non-predictive condition, less time has to pass. This difference in time between the predictive and the non-predictive case on the objective time axis should reflect how much earlier a participant perceives the predictive character subjectively (compare subjective time axis); as the predictive character is perceived earlier, the test interval should get subjectively shorter. Hence, to perceive the test interval to be as long as a standard interval, it has to be objectively lengthened as compared to the non-predictive case. This lengthening should correspond to the temporal benefit in perceiving a predictable character and should be expressed by a difference in PSEs—namely a rightward shift of the psychometric function for the predictive case.

The right part of Figure 7 shows a schematic of a virtual subject's responses in both experiment conditions and displays the resulting psychometric function. Again, the predictive condition is shown in the upper part and the non-predictive condition in the middle. Both conditions are again split into an objective and a subjective time axis. A plot of both psychometric functions is displayed in the lower right. Colored squares represent one single trial each. Squares give subjective responses of this virtual single participant to the question of whether the last interval was longer or shorter than the previous intervals. Red squares indicate the judgment "too short," while blue squares indicate "too long" responses. Each vertical column of squares represents one specific length of the varying test intervals. The rows reflect individual trials of a given test-interval length.

The sigmoidal curves represent the resulting psychometric function for each condition. If the last interval is very short (very left column), it is quite likely to be judged as "too short" (red squares). If the last interval is very long (very right column), it is quite likely to be judged as "too long" (blue squares). If the last interval is of intermediate duration, it is harder for the participant to make a decision. Therefore, columns in the middle—representing intermediate interval lengths—are mixed with red and blue squares, representing both “too short” and “too long” judgments. Again, as each interval length of the test interval is repeatedly used in several trials, several rows of squares are depicted.

The sigmoidal psychometric function expresses the percentage of “too long” answers as a function of test interval duration. Columns on the very left with short test intervals show only “too short” answers. Hence the curve is close to 0%. On the right side, where test intervals are very long, all answers are “too long.” Hence, the curve approximates 100%. The intermediate area of the curve has a steep slope. From this curve, the 50% point (PSE) of “too long” answers is taken. This value does not necessarily exist as an interval length that was really presented to the subject. This value is calculated from the curve and not measured directly. It might have a value between two test-interval lengths.

The psychometric functions of the two conditions are of the same shape but are shifted on the x-axis. For the predictive condition (Figure 7, greenish color), the curve is located relatively to the right, whereas for the non-predictive condition, the curve is shifted to the left (greyish color). This is due to a change from “too short” answers to “too long” answers at shorter test intervals for the non-predictive condition. When both curves of this (theoretical) subject are collected in one graph, the relative shift of the curves becomes obvious (Figure 7, lower part).

As mentioned before, we have hypothesized predictable letters to be perceived earlier. We expected a relative shift of PSEs. However, this method cannot determine whether a predictable character is in fact represented earlier than an unpredictable character or, alternatively, if the violation of a prediction in the non-predictive case leads to a “delayed” representation. Our MEG experiment is designed to enable us to address this question, as detailed below.

Eye Tracking

Eye movements were monitored using an eye-tracking system from Arrington Research (Monocular PC-60 integrator system, PCI framengrabber and ViewPoint eye tracking software). The IR-camera system was positioned in front of the right eye while two IR lamps illuminated the recorded eye (see Figure 9). Sampling rate was 50 Hz. Eye tracking data of one single trial are illustrated in Figure 8.

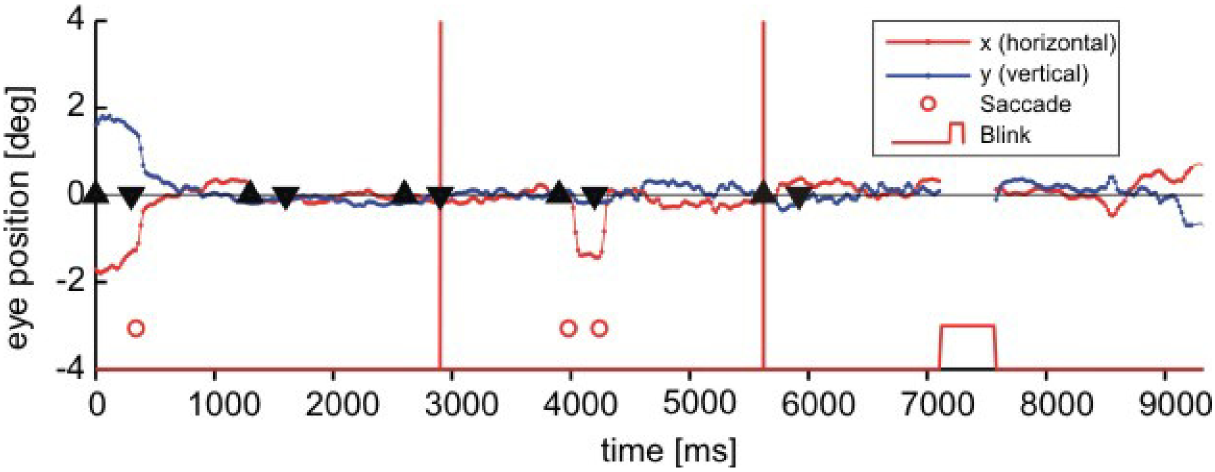

Figure 8: Eye tracking data from one single trial.
Blue line visualizes vertical eye movements; red line tracks horizontal eye movements. Vertical red bars mark the beginning and end of the phase, in which eye movements and blinks led to the exclusion of a trial (as in this example). Triangle pointing up indicates characters' appearance and triangle pointing down their disappearance. Saccades appeared around 300 ms and 4200 ms and a blink around 7500 ms. Red and blue eye-tracking lines are "cut out" in case of a blink artifact.

Although one could argue that blinks and saccades should be equally distributed across conditions and therefore should not cause a systematic difference in our results, we excluded trials with eye movements/blinks in a certain time epoch from our analysis. This is because they might mask the hypothesized timing-effect and therefore more trials per subject might be required to reveal the hypothesized effect. For a more complete discussion on this issue, please refer to the discussion section.
Eye movements were analyzed for a critical time window at the end of each trial, which lasted from the end of the antepenultimate letter till the end of sequence. If a subject performed either a blink or an eye movement during this time window, the respective trial was labelled invalid and was not taken into account for further analyses. Blinks and saccades were defined by eye position and velocity[10]. Subjects were instructed not to blink or perform saccades while the stimulus was presented. During the response phase, however, participants were allowed to blink (and move their eyes).

[10] Detection threshold for saccades was 15°/s.

IR-Lab Setup & MEG Setup

Setups: We used two different setups with the same software to realize experiments. Setup was situated either in the IR laboratory (therefore named IR-setup) or at the MEG center. Stimuli were generated with Cogent Graphics toolbox 1.29 (http://www.vislab.ucl.ac.uk/cogent.php) and MATLAB v7.7.0 R2008b (MEG lab) and MATLAB v7.5.0 R2007b (IR lab) both from “The MathWorks.” At the MEG setup, stimuli were presented via back-projection screen, with spatial resolution of 1024 by 768 pixels and at 60 Hz refresh rate. At the IR-lab setup, stimuli were presented on a CRT screen with a spatial resolution of 1280 by 1024 pixels and at 100 Hz refresh rate. During experiments at the IR lab, a head and chin rest stabilized the subjects’ head (see Figure 9). For both setups, screen width was 40 cm, with 57 cm of viewing distance. All experiments were performed in a darkened room to enhance stimulus visibility. A video-based dark-pupil eye-tracking technique was engaged in both setups to register participants’ eye movements. See above for further details on eye tracking.

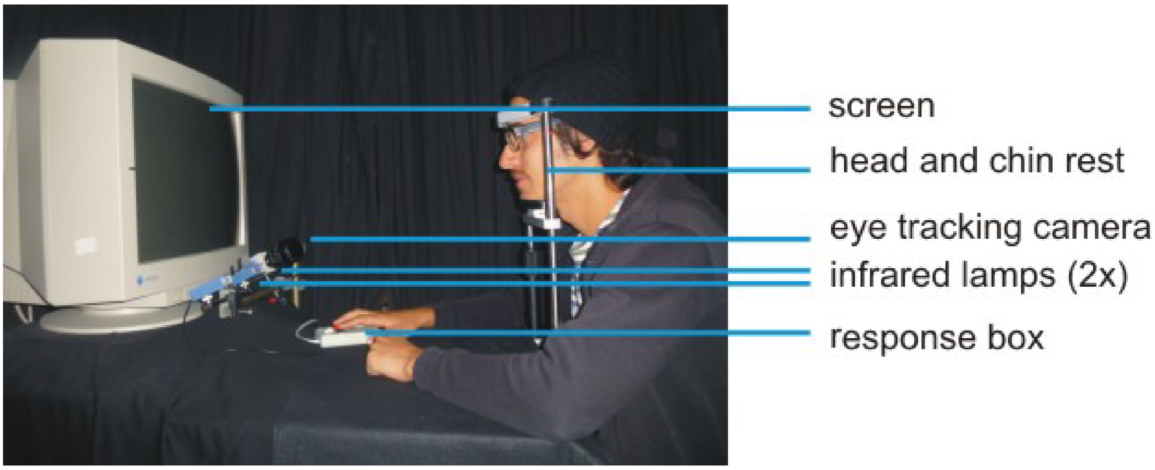

Figure 9: Photo of setup at the IR laboratory.

The subject sat in front of a CRT screen (100Hz) at 57 cm distance. The subject’s head was stabilized by a chin-headrest. Two IR lamps were installed in front of the subject and an eye-tracking camera was placed right beside the screen. A response box with two buttons was provided to register the subjects’ manual answers.

The MEG setup underwent a minor change compared to IR-lab setup due to the physical restrictions of the MEG recoding environment. Projection was restricted to 60 Hz due to the technical limitations of the video projector. The subjects sat in the MEG scanner in a magnetically shielded room (see Figure 10).

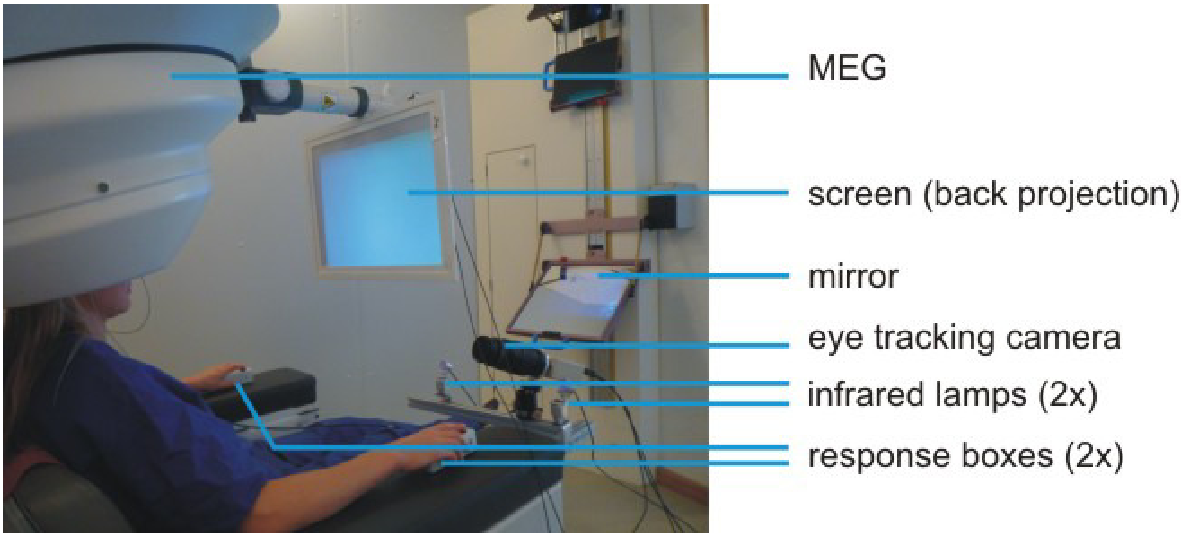

Figure 10: Photo of setup at MEG facilities.
The subject sat in front of a back-projection screen (60Hz) at 57 cm distance. The subject's head was positioned in an MEG "cap" and stabilized by foam paddings. Two IR lamps and one eye-tracking camera were installed in front of the subject on a bar. Two MEG-compatible response boxes were provided to answer through button presses.

They were asked to avoid movements. Their head was stabilized by foam pads. The experiment lasted approximately 1 hour 30 minutes for both pretest and main experiments. Neuromagnetic data were acquired using a whole-head MEG system (275 axial gradiometers, CTF MEG by MISL, Port Coquitlam, BC, Canada).

Magnetoencephalography

MEG is a method to infer electrical currents in the brain at a submillisecond ($<10^{-3}$ s) temporal accuracy. Electrical currents in the brain generate weak magnetic fields, which are recorded with sensitive detectors during MEG study (Toga & Mazziotta, 2002).

External superconducting quantum interference device (SQUID) sensors are used to record evoked potentials on brain surface. Evoked potentials are the summed electrical activity of thousands of cells (Kandel et al., 2011), either evoked by sensory perception or related to cognitive processes. A visual-evoked potential (VEP) is elicited by a visual stimulation. Components of evoked potentials are termed according to polarity (N=negative, P=positive) and latency (in milliseconds from stimulus onset). N100, for example, is the negative component of a typical VEP assessed by electroencephalography (EEG), arising 100 ms after stimulus onset with negative polarity (Stöhr & Bach, 2005). Corresponding signals can likewise be traced by MEG—such as the M100, which is the MEG version of the N100 (EEG).

Eye movements and blinks may cause MEG artifacts (Toga & Mazziotta, 2002). Therefore, it is important to monitor them and to instruct subjects to avoid them during stimulus epoch in our MEG study.

With the collection of MEG data, we aimed to investigate electrophysiological correlates of sensory predictions. Since MEG is not retarded by hemodynamic delays, it also provides high temporal accuracy. High temporal resolution is important when studying time perception. This makes MEG an adequate method to search for neuronal correlates of time perception in our study.

Experiment Procedure

Subjects attended at least two training sessions—the first without eye tracking and the second with eye tracking. In the first experiment design, these trainings are followed directly by the experiment. For the second and all subsequent designs, training is followed by a pretest, before the main experiment starts.

Pretest

We conducted pretests for all but the very first design to roughly estimate each subject's individual PSE for our sequential temporal interval judgment (not distinguishing predictive and non-predictive cases). This estimate allowed a more fine-grained sampling of the subjects' psychometric curve around this PSE during the main experiment using the method of constant stimuli (MCS) (Wirxel & Lindner, 2012).

For pretest, the length of the varying "test interval" was governed by an adaptive staircase procedure (Parameter Estimation by Sequential Testing or PEST, targeting the PSE (Taylor & Creelman, 1967)). Hence, test interval lengths depended purely on the subject's temporal judgment. This allows the experimenter to ensure that each subject is tested in his or her individual range of time perception. The experimenter defines only start and target values. (For these and further parameters, see appendix "tabular overview of experiment designs"). The staircase procedure calculates all subsequent test values automatically, based on a subject's temporal judgments. Most

pretest designs[11] included three strategies, each with a different target value—25%, 50%, and 75%. For example, the 25% strategy aims to reveal the PEST value, in which 25% of subjects' temporal judgements are "too long." For slight changes in pretest design, see appendix "tabular overview of experiment designs" for designs.
Furthermore, we calculated the just noticeable difference (JND) as a parameter for our exclusion criteria. It reflects the smallest detectable difference of two sensory stimuli (here interval durations). The JND is half of the difference of the 75% value from the 25% value [(75% − 25%)/2] of the psychometric function. Since the psychometric function is symmetric, it is equal to the difference in test interval duration between the 75% and the 50% value of the psychometric function. If her/his JND fell within the JND criterion, a subject was allowed to proceed with the main experiment. The pretest consisted of 90 valid trials. From the pretest data, a PSE value (pre-PSE) was calculated. Pre-PSE was then fed as a parameter/"central value" into the main experiment. To calculate the pre-PSE value, the two conditions, P and NP, were taken together. A probit analysis was performed on this merged dataset to calculate the pre-PSE.

Main Experiment

The main experiment was based upon another test strategy, known as MCS. In contrast to a staircase procedure, the varying test interval did not depend on the subject's temporal judgments. Instead, interval length was fixed to 11 specific values (11 different test-interval lengths). These fixed values for the test interval were calculated before the main experiment started. Pre-PSE was taken as central value (rounded to the actual time of presentation of screen frames in the respective setup). For IR setup, it was 10 ms, while for MEG setup it was a multiple of 16.667 ms in whole numbers (0; 17; 33; 50; 83; 100). For further details, please see "MEG" section. All test intervals are calculated according to this. Test intervals are clustered around the central value and equally distributed up to 250 ms up and 250 ms down, in 50 ms steps.
Each test interval length was tested10 times for each interval length. This summed up to 110 trials for each condition (11 interval length x 10 trials). Thus, 110 predictive trials and 110 non-predictive trials summed up to 220 trials all together. In addition,

[11] For Design #02, the first pretest version included less strategies (one 50% -strategy for each condition). For further details see appendix 'tabular overview of experimental designs'.

each invalid trial had to be repeated. In cases where the alphabetical question was answered incorrectly, the trial became invalid. Likewise, trials became invalid if eye movements/blinks took place during the defined critical period, but only in those experiment designs that incorporated an online control of eye movements. To collect the required amount of trials for each condition, invalid trials were repeated immediately, given the required set of parameters.

The alphabetical question was used as a tool to make subjects focus on the content of the characters and as a control. If subjects were allowed to purely focus on the rhythm of appearing and disappearing characters, they might not have taken characters' content into account. This study was based on the difference between a predictable condition (predictable last character) and a non-predictable condition (non-predictable last character). Therefore, it was crucial to make subjects focus on predictable information and thus on the content of the stream of characters. This was why the alphabetical question was introduced. In case of an incorrect answer to this question, the trial became invalid. In addition, feedback on alphabetical performance was provided to keep subjects motivated. In case of online-control of eye movements, feedback on eye movements was provided to help subjects improve their eye movement performance. Trials including eye movements during a defined phase were labelled invalid as well, as described earlier.

Subjects got five seconds to answer each question. If they exceeded this time limit, the trial became invalid. Hence, if a subject could not see what happened on the screen, she or he just had to wait for five seconds. This could also be used to relax the eyes a little; if a subject wanted, she or he closed the eyes for the length of one or two trials and then restarted with the next trial.

Exclusion and Inclusion Criteria

Since the experiment required precise timing, we apply rather strict exclusion criteria to both subjects and trials (Wirxel & Lindner, 2012).

For subjects, we applied these criteria:

- adult human subjects
- no diagnosed neurological or psychiatric disturbances
- normal or corrected-to-normal vision (according to subjects' statements)
- raised with Latin alphabet in first language

In addition, for the MEG setup:

- MEG: no metal in body or in clothes (to avoid artifacts)

For data analysis, we applied these criteria for trials:

- Coverage of “full” psychometric function
- JND criterion
- Reliable fit of psychometric function ($p<0.05$)
- Exclusion of single trials with blink or saccades during the pre-defined time period (see “eye tracking” section)
- 100 or more valid trials left after trial exclusion (for those experiments with post-hoc analysis of eye movement data)

In general, a steep slope reflects a high precision of temporal order judgment and is reflected by a small value for the JND. We calculated JND for pretest and main experiment in two different ways and for two different purposes. For the pretest, we calculated a combined JND-value pooling data of both experiment conditions. For the main experiment, we calculated separate JNDs for both conditions (Wirxel & Lindner, 2012). This was for two reasons. First, for the pretest, we took JND to exclude those participants who were likely to require far more than 220 trials in the main experiment to allow an adequate analysis of the data. A subject who is generally very uncertain about time judgment will show a shallow curve and a high JND value. To come up with a good estimate for the pretest PSE (pre-PSE), the function either needs to be steep, or, alternatively, more data points are required. If the pre-PSE is unreliable, due to a shallow curve, the estimate of the pre-PSE might be considerably wrong. This “wrong” value would then be used as a central test value for the main experiment. Therefore, the psychometric function would not be ideally sampled in the main experiment. Second, for the main experiment, JND values were helpful for interpreting PSE values. If PSE values of predictive and non-predictive conditions differed significantly, while related JND values of the two conditions did *not* differ significantly, the difference in PSE values could *not* be explained by differences in JND values. Therefore, it is important to have a look at these values as well.

Specific Experiment Designs, Results, and Implications

To establish a paradigm suitable to test our hypotheses, we needed to develop and test several experiment designs. Finally, we came up with a promising design, which we also transferred to MEG to get insights into the electrophysiological underpinnings of the phenomena under investigation. The designs were the following.

Design #01—Pilot study

Design #01: Concept

The first paradigm (#01) was designed according to the basic concept, as described before. We realized the variation of test interval length by implementing an adaptive staircase procedure called PEST (Parameter Estimation by Sequential Testing) (Taylor & Creelman, 1967). PEST is a method for psychophysical testing, designed to reveal detection thresholds. The duration of test intervals varied depending on the subject's previous time judgments. Only for the first trial of each strategy were start values set by the experimenter. There were six PEST strategies—three each for the predictive and non-predictive conditions. Start values were the same across both conditions: 1500 ms, 1200 ms, and 300 ms. Initial step sizes were −320, −320, and 320 for each condition. WALD constant was 1. For further parameters, see appendix "tabular overview of experiment designs." Standard intervals were displayed for 1,000 ms and characters for 300 ms. The experiment consisted of 180 trials in total—90 non-predictive (NP) and 90 predictive (P). The 90 trials of each condition were split into three PEST strategies, each of which had a different target value—namely 25%, 50%, or 75% "too long" responses. This was a tool to make sure that a broad range of the individual's psychometric function was covered by sampling points. The control task was not controlled online; therefore, no feedback was given. We excluded trials with "incorrect" answers to this task post hoc, which caused a decrease in the number of evaluable trials. The likelihood for the last character to be predictive was 50%.

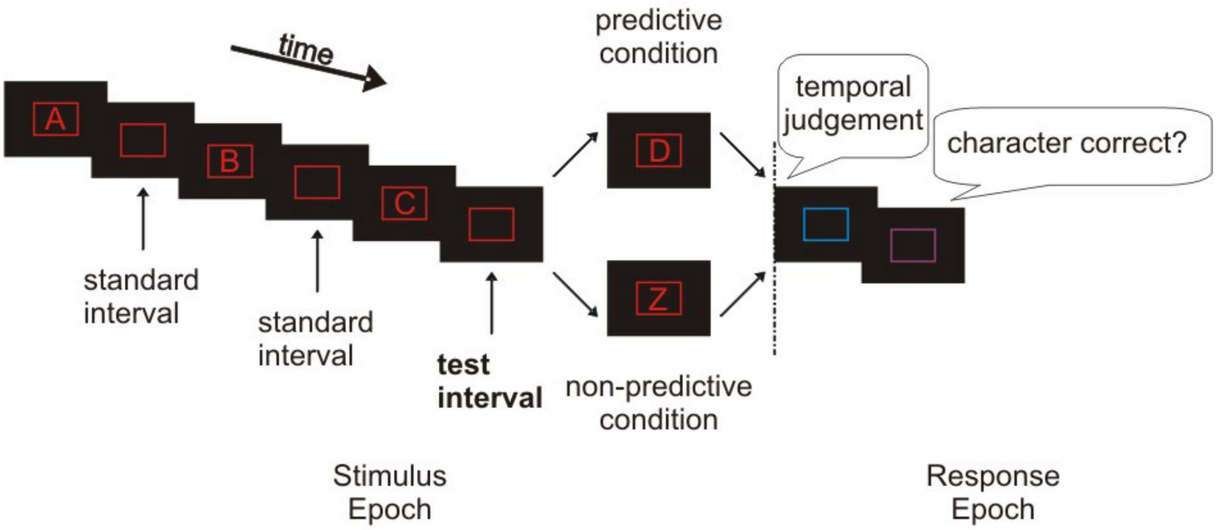

Figure 11: Experiment task design (#01)

The first design exactly followed the basic idea. Standard intervals were displayed for 1,000 ms and characters for 300 ms. Test interval length varied, depending on previous time judgments of the subject. For the predictive condition, the last character followed the alphabetical order (upper part, "D"). For the non-predictive condition, the last character did not follow the alphabetical order (lower part, "Z"). The subjects judged the length of the test interval compared to the previous standard intervals (temporal judgment, blue rectangle). Their second task was to indicate whether the last character was "correct" or "incorrect" according to alphabetical order.

Design #01: Results and Implications

We conducted the experiment 12 times (n=12) to get pilot data. Prior to evaluating pilot data, we excluded all trials with incorrect answers to the alphabetical order-question.

The hypothesized effect of a difference between predictive and non-predictive conditions was not revealed in the data of Design #01. This was a negative result. In detail, PSE values of the two conditions did not show a statistical difference (see Figure 6, left). The intra-individual difference, i.e. the difference occurring within an individual, was termed as paired difference: the PSE value of all predictive trials of one individual was taken and the PSE value of all non-predictive trials of the same individual was subtracted (see Figure 6, center). We expected the paired difference of individuals to cluster below 80 ms. Processing delays are known (from animal research) to take 80–100 ms, depending on the investigated visual area (see introduction).

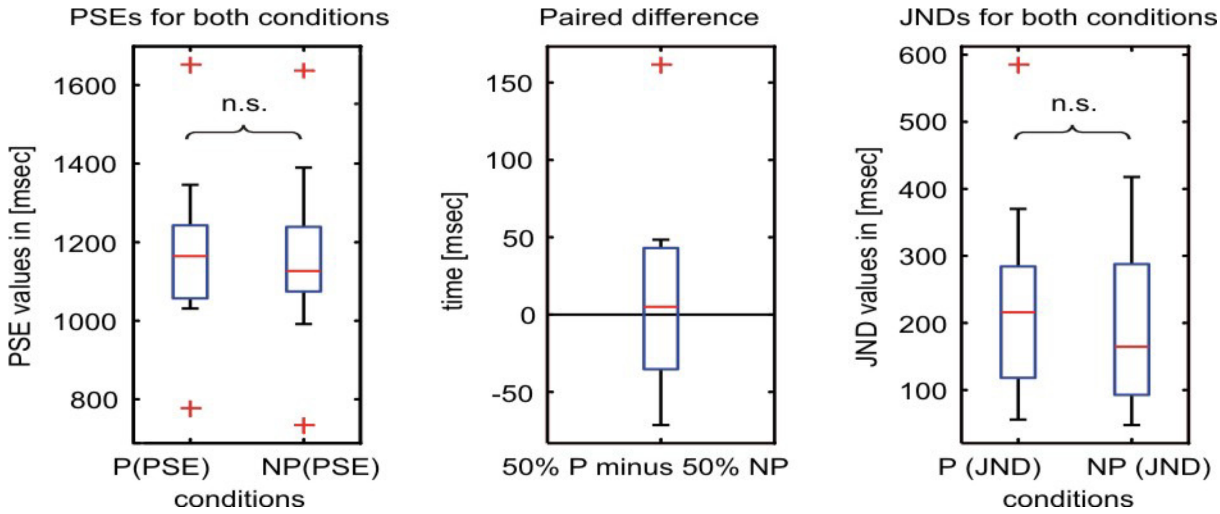

Figure 12: Data from paradigm #01.

On the left, PSEs of both conditions show no significant difference (p= 0.586, two-tailed paired t-test, alpha = 0.05). The hypothesized effect of a difference of conditions is absent. The same interpretation may be taken from the central boxplot. Paired differences cluster around zero. Median value (red) is slightly above zero at 5 ms (mean 10.4 ms). JNDs (right) show a broad variance and do not differ significantly (p=0.1482, two-tailed paired t-test, alpha = 0.05). (n=12; All boxplots show quartiles in blue and median in red. Whiskers give 1.5 times the interquartile range; outliers[12] are displayed with a red + sign. Incorrect trials according to the alphabetical order question are excluded from the data set).

Therefore, we considered 80 ms as the "upper" limit of what needs to be compensated for. In addition, we did not necessarily expect a full compensation for neuronal processing delay on perceptual level. Different mechanisms could contribute. Therefore, we expected a value below 80 ms for the paired difference, which should be significant above 0 ms. For this design (#01), paired differences clustered around zero. Since the paired difference is just another way to visualize the difference of P-PSE and NP-PSE, it was of no surprise to see a median with 10.4 ms close to zero. JNDs showed large values, mostly above 100 ms, while the upper quartile was close to 300 ms (see Figure 6, right). High JND values indicated a high uncertainty in the individuals' time judgments. JNDs of the two conditions did not differ significantly.

[12] Outliers: are defined as values > Q3+ 1.5 (Q3-Q1) or < Q1 – 1.5 (Q3-Q1), where Q1 is the 25th percentile and Q3 is the 75th percentile.

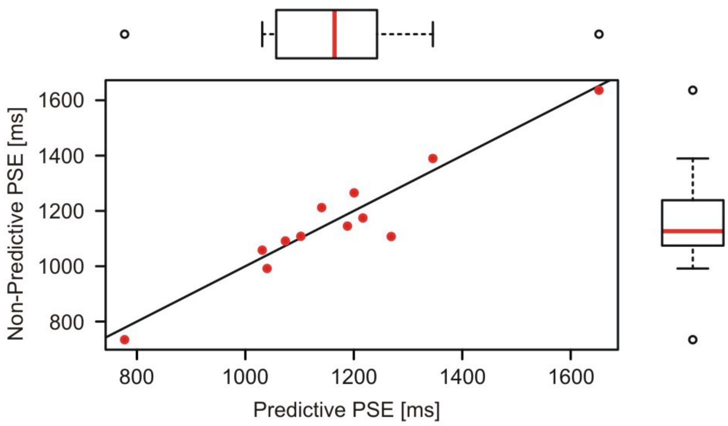

Figure 13: **PSEs from predictive and non-predictive conditions (Design #01, n=12).**
The diagonal line indicates the point at which PSE values for the two conditions (P, NP) equal each other. Every dot below the diagonal represents a subject whose predictive PSE was longer than the non-predictive PSE. Six out of 12 subjects showed larger PSE values in the P condition (data below line).
There was no significant shift of the PSE toward larger values for predictable sequences (median shift: 5ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Median values (central bars) are 1164 ms (P) and 1127 ms (NP). The empty black circles mark the outliers.

Figure 13 shows another way to plot the data. The scatter-box plot combination provides box plots for both the predictive (x-axis) and the non-predictive (y-axis) condition. The diagonal line indicates the point at which PSE values for the two conditions are equal. Data clustered around the diagonal line. Half of them were above and half below the line. This is in agreement with the central plot ("paired difference") shown in Figure 12.

Pilot data revealed the typical range for test intervals to be between 300 ms and 2,000 ms. Surprisingly, PSEs showed a tendency to attain values above 1,000 ms (see Figure 12, left). A naïve reader/person would perhaps expect the PSE values to cluster around the standard interval length of 1.000 ms. This phenomenon is unlikely to have been caused by the experiment test strategy (i.e. PEST) due to its adaptive nature: "Test interval" length purely depended on the subject's previous temporal judgements. There was no tendency to present (on average) longer "test intervals," which might have led to higher PSE values—e.g. through "temporal adaptation." Therefore, the strategy did likely not cause the "above-1.000 ms" phenomenon. Also, in the present study, we have focused on differences of predictive vs. non-predictive condition. The tendency toward larger values was present for both conditions and hence should cancel out in their relative comparison. In addition, for the main experiment, we did not apply the PEST strategy, but the MCS. Hence, “test-interval” length did not depend on the subject’s previous answers during the main experiment. Since this phenomenon can be observed across all subjects, it seems to be inherent to people’s temporal judgment, given the experiment conditions. Potential explanations are Vierodt’s law and time-order errors—in particular a negative time-order error. For more details, please refer to the general discussion.

Because of the negative result, an adjustment was required. The adjustment aimed to choose the most suitable test-interval length. Hence, we changed from PEST to MCS and split the experiment into two parts—the pretest and the main experiment. We based the pretest on a PEST strategy and the main experiment on an MCS. By designing a pretest, the subject’s individual range of the psychometric function could be roughly estimated and used to tailor the main experiment to every single subject’s range of time perception. We implemented this idea in Paradigm #02.

To illustrate this change in data collection-strategies, I plotted the data of a single subject from Design #01 and of a subject from Design #02, including pretest and main experiment (Figure 14, Figure 15, Figure 16, Figure 17). Figure 14 and Figure 15 show how the data was collected in Design #01 by applying PEST strategy. Both figures illustrate the course of the experiment and the staircase procedure.

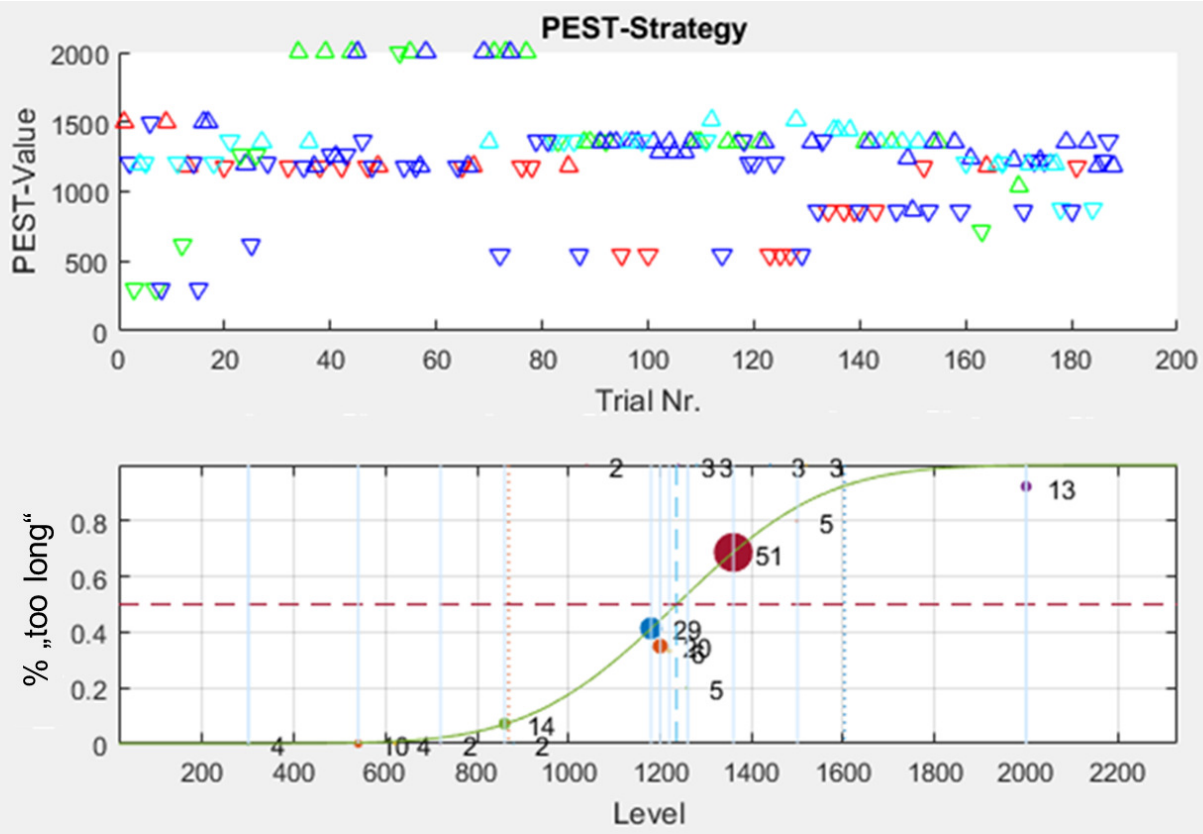

Figure 14: Raw data from temporal judgement task of subject no. 10, Design #01 (PEST strategy).
Upper part **shows every temporal judgement given by the participant. Rectangles give duration of test intervals in ms (y-axis, PEST value) and respective trial number (x-axis). Rectangles pointing up indicate the test interval being judged as "too long," whereas rectangles pointing down indicate the test interval being judged as "too short." Standard interval length was 1,000 ms. Rectangles pointing down for PEST values below 1,000 ms illustrate a correct temporal judgement. Rectangles pointing up for PEST values above 1,000 ms illustrate correct temporal judgements as well. The colors of rectangles indicate different PEST strategies. Three strategies were applied for each condition (3*P, 3*NP).** ***Lower part*** **shows the psychophysical function of temporal judgements. X-axis gives the proportion of temporal judgements judged as "too long." A proportion of 0.5 indicates that the participant judged in 50% of the cases for the respective PEST value ("level" on x-axis) and is also known as the point of subjective equivalence (PSE). Since it derives from the pretest, it is termed pre-PSE.**
Data include all trials (invalid trials are not eliminated yet).

For Design #01, six PEST strategies were intermingled, three strategies for predictive and three for non-predictive condition. For start values and initial step sizes, see appendix "tabular overview of experiment designs." All subsequent test values were calculated automatically based on the subject's temporal judgments. To increase the visibility of a single strategy (of staircase procedure), the data from one single strategy is shown in Figure 15.

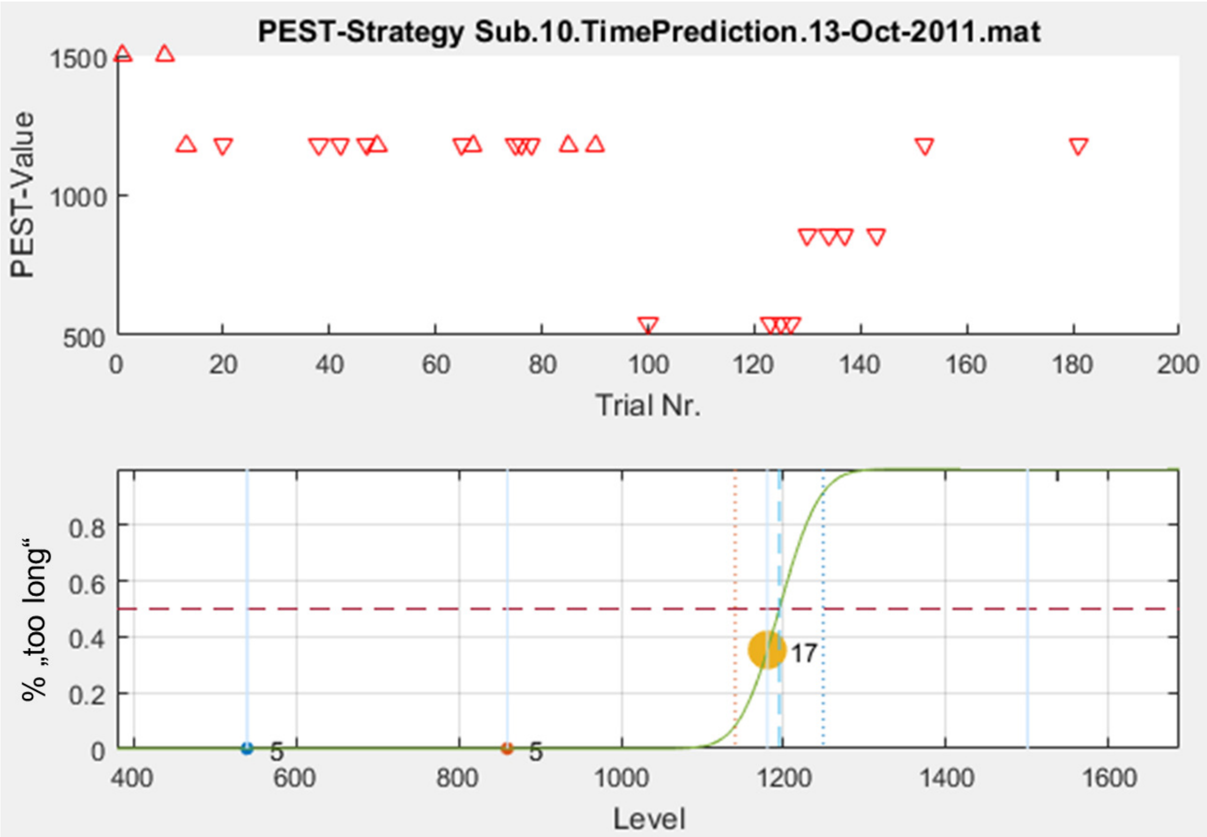

Figure 15: Raw data from temporal judgement task of one *single* PEST strategy of subject no. 10, Design #01.

***Upper part* shows every temporal judgement given by the participant for the respective PEST strategy. Rectangles give the duration of test intervals in ms (y-axis, PEST value) and respective trial number (x-axis). Rectangles pointing up indicate the test interval being judged as "too long," whereas rectangles pointing down indicate the test interval being judged as "too short." Standard interval length was 1,000 ms. Rectangles pointing down for PEST values below 1,000 ms illustrate a correct temporal judgement. Rectangles pointing up for PEST values above 1,000 ms illustrate correct temporal judgements, as well. The rectangles of one strategy alone show a staircase pattern. This pattern is a consequence of the PEST procedure. *Lower part* shows the psychophysical function of temporal judgements for the respective PEST strategy. X-axis gives the proportion of correct answers for the temporal judgement. A proportion of 0.5 indicates the participant made a correct temporal judgement in 50% of the cases for the respective PEST value ("level" on x-axis). Data includes all trials of respective PEST strategy (invalid trials are not eliminated yet).**

For Design #02, the pretest and main experiment were established (see section "Design #02—Full Experiment"). Figure 16 and Figure 17 illustrate how data was collected in Design #02, by applying PEST strategy for pretest and MCS for main experiment. They illustrate the course of experiment. For a general description of the split into pretest and main experiment, see "experiment procedure" section.

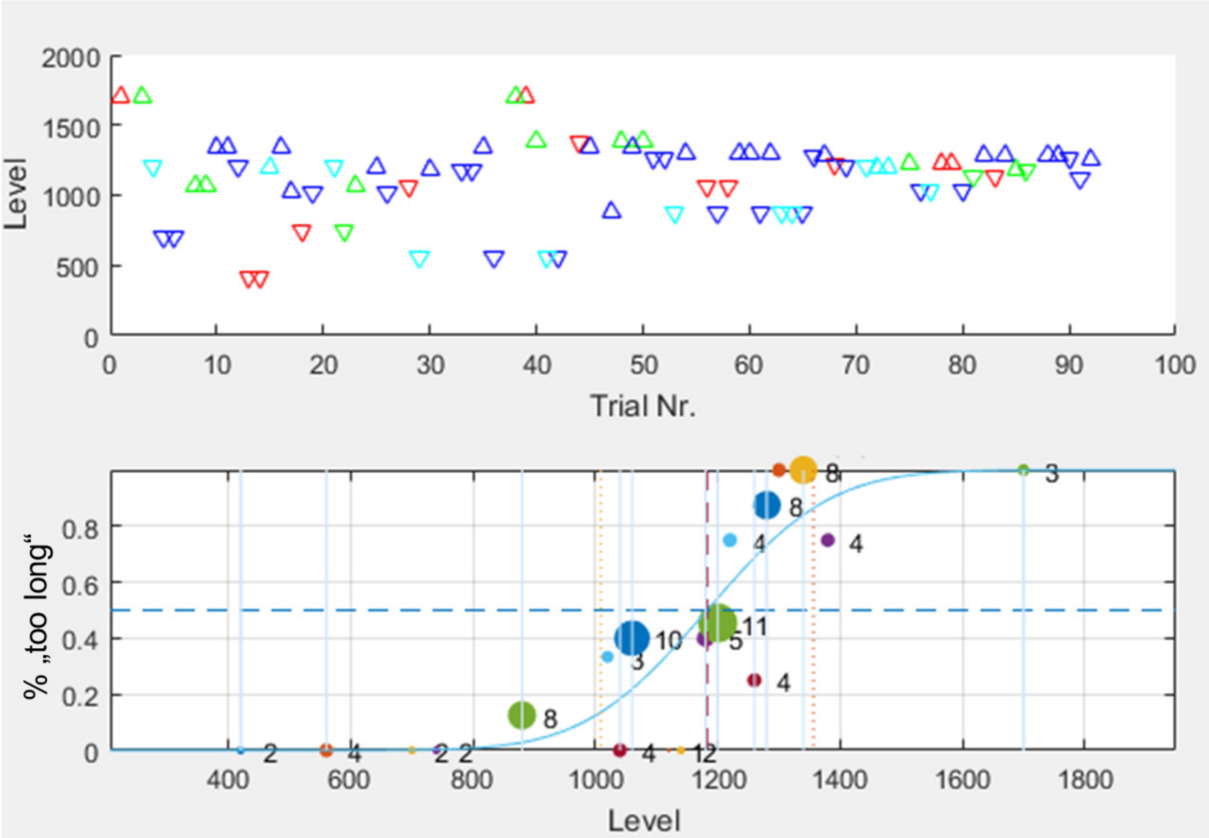

Figure 16: Raw data from temporal judgement task of pretest of subject no09, Design #02 (PEST strategy).
Upper part **shows every temporal judgement given by the participant. For details on rectangles, see caption of Figure 14. Colors of rectangles indicate different PEST strategies. Three strategies were applied for each condition (3*P, 3*NP).** ***Lower part*** **illustrates all trials of six PEST strategies (3*P and 3*NP). Pretest PSE for this dataset is 1,183 ms.**

Data collection in Design #02 is illustrated in Figure 16 and Figure 17. A split into pretest and main experiment allowed us to collect data in the range that was relevant for an individual subject. For further details on this split, please see section “Design #02—Concept.” Pre-PSE (1,183 ms) was rounded off to 1,180 ms and fed as the central value into the main experiment (MCS). The JND of pretest data was calculated and fulfilled the JND criterion of <400 (JND-value was 107). Hence, the subject was allowed to proceed.

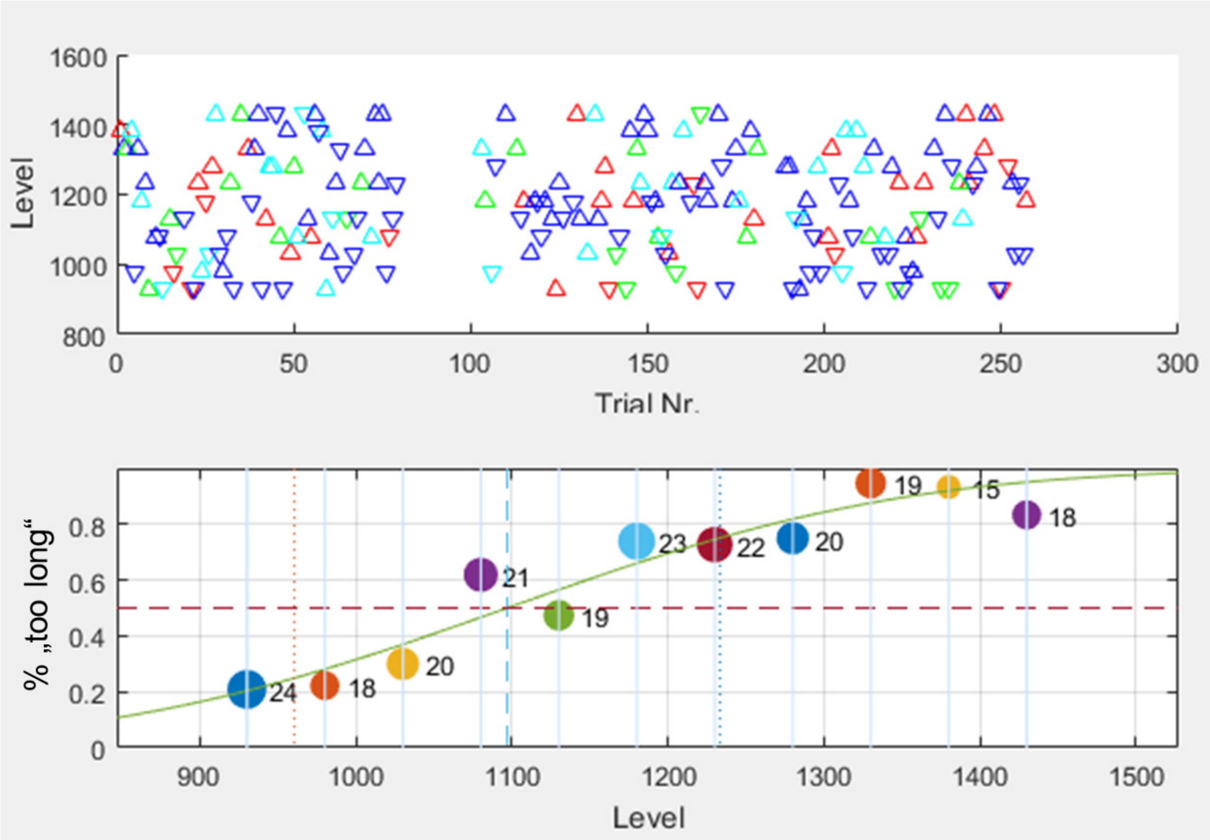

Figure 17: Raw data from temporal judgement task of P and NP-trials from main experiment of subject no. 09, Design #02 (MCS).

Upper part **shows every temporal judgement given by the participant. For details on rectangles, see caption of Figure 14. The "gap" around trial no. 100 is due to a pause during the experiment.**

Lower part **gives the psychometrical function. PSE is 1,096 ms (P and NP taken together).**

Test values for the main experiment sample clustered around the central value 250 ms up and down—in this particular case 1,180 ms ± 250 ms (930 ms to 1,430 ms). The psychophysical function of the main experiment seems to be shallow at the first glimpse. It seems so because range of test values was much broader in the pretest than in the main experiment. Please compare the "level" values for Figure 16 and Figure 17. By applying a pretest based on a dynamic PEST strategy, we were able to come up with a good estimate for a central value for the main experiment. Around this central value, we could apply a fine-grained sampling during the main experiment by using the MCS.

Design #02—Full Experiment

Design #02: Concept

The substantial novelties of this design were the incorporation of the pretest and the provision of feedback regarding the control task (alphabetical order task). The subject's performance of the control task was immediately evaluated (online control): If the answer to this task was incorrect, the trial became invalid. Written feedback was given, as shown in Figure 18, by the word "text" during feedback epoch. In detail feedback was either "*Alphabet falsch beantwortet*" (alphabet answered incorrectly) or "*Alphabet richtig beantwortet*" (alphabet answered correctly). Feedback was beneficial to keep subjects motivated and trained them to focus on character content. This was of importance, as described above.

The novel pretest required additional time and effort from our subjects. To charge the subject as little as possible with this extra-task, we designed it to be slightly different from the PEST-based experiment #01. Compared to Design #01, we reduced the number of trails. Soon, we realized that the first pretest design needed adjustments to increase its effectiveness. The first pretest design was used for subjects no. 1–8. Most importantly, the first pretest was designed with only one strategy for each condition (50% strategy for predictive condition, and another 50% strategy for non-predictive condition). This led to the following problem: If a subject was only *guessing* (giving 50% of the answers as "too long" based on guessing and not on really judging time), a 50%-strategy may converge toward its 50% target-value. Therefore, we designed a new pretest, including 25%-strategy and 75%-strategy for each condition. In addition, Wald's constant was reduced from 1.0 to 0.75, which makes the strategy more agile. For these and further parameters, see appendix "tabular overview of experiment designs."

Taken together, these changes allowed us to modify the JND criterion to allow subjects to participate in the main experiment. Details of change in the JND criterion can be found in section "Design #02—results and implications." The later ("new") pretest—which was used for subject no. 09 and subsequent subjects—is described in the following. Modified pretest consisted of 90 trials and the changed start values were 1,200 ms, 1,700 ms, and 700 ms. Initial step sizes were 1,280, 1,280, and −1,280. Wald-constant was reduced to 0.75. (For these and further parameters, see appendix "tabular overview of experiment designs"). Those values were the same across both conditions (as before). For both conditions, (P&NP), only one PEST strategy was applied with a target value of 50% “too long” responses (i.e. the PSE). Standard intervals were displayed for 1,000 ms and characters for 300 ms. This was consistent with the main experiment as well as with Design #01. Due to the limited number of pretest trials, the PSE of the pretest (pre-PSE) was calculated across both conditions. Afterward, we fed the estimate of the PSE into the main experiment as a central value of the MCS procedure.

In detail, the calculation of pretest PSE and test-interval lengths for the main experiment was as follows: Calculating pretest PSE across both conditions (P&NP) was feasible. This was because we expected the difference in time perception between both conditions to be in the range of 30–80 ms, as it should partially compensate for a processing delay of 80–100 ms. The aim of the pretest was to get a rough estimate of each subject’s individual range of time perception and, in particular, of the average PSE. During the main experiment, a more fine-grained sampling around the PSE should be realized using the MCS. We defined the range of MCS values so that they clustered around the pretest PSE (pre-PSE). Pre-PSE served for the calculation of all test intervals for the respective subject. A total of 11 test intervals were defined within 250 ms above and 250 ms below the central value (pre-PSE). Specifically, the pre-PSE[13] was rounded to 10ms values and served as the central value and the remaining MCS values were chosen in 50ms steps, i.e. pre-PSE-250 ms; pre-PSE-200 ms; pre-PSE-150 ms; pre-PSE-100 ms; pre-PSE-50 ms; pre-PSE+50 ms; pre-PSE+100 ms; pre-PSE+150 ms; pre-PSE+200 ms; pre-PSE+250 ms. Hence, a range of 500 ms for the test interval length was covered in the main experiment. Whether the pre-PSE was slightly imprecise due to the combined pre-PSE calculation across both conditions should not matter too much as we expected the difference between both conditions to be 50–100 ms at most and since this difference should be covered by the 500ms interval range in the main experiment.

Actually, it should be advantageous to calculate the pre-PSE across the two conditions. Since the pre-PSE was based on a mixture of behavioral data from predictive and non-predictive conditions, it could be used to calculate the test interval

[13] Pre-PSE was rounded to 10ms values before it was fed as the central value in the main experiment. We chose to round to 10ms values because of the 100Hz refresh rate of the display.

lengths for both conditions for the main experiment. This was of importance, because test intervals had to be equally long across the two conditions during the main experiment. Otherwise, one might argue that a difference found in behavioral data might have been induced by this systematic difference of test interval lengths based on different pre-PSE values for the predictive and non-predictive conditions. To summarize, for pretest data, calculating PSE across both conditions—P and NP—was feasible. Clearly, we did not pool data across conditions in the main experiment because determining the difference of P and NP in terms of the PSEs was the aim of this investigation. Calculation of pretest PSEs across both conditions was an adequate means to keep the overall trial number low (pretest) and to guarantee comparability (with the main experiment). The experiment task design otherwise did not differ between the pretest and the main experiment. Only the underlying strategies for the calculation and presentation of the test intervals differed. Therefore, the experiment task design (#02) in Figure 18 is representative for both experiment subparts. The change from Design #01 to Design #02 was mainly at the level of experiment strategies (PEST vs. MCS); these changes are not visible in the illustration of the task design. This illustration only differs with respect to the novel feedback stimulus for Task 2.

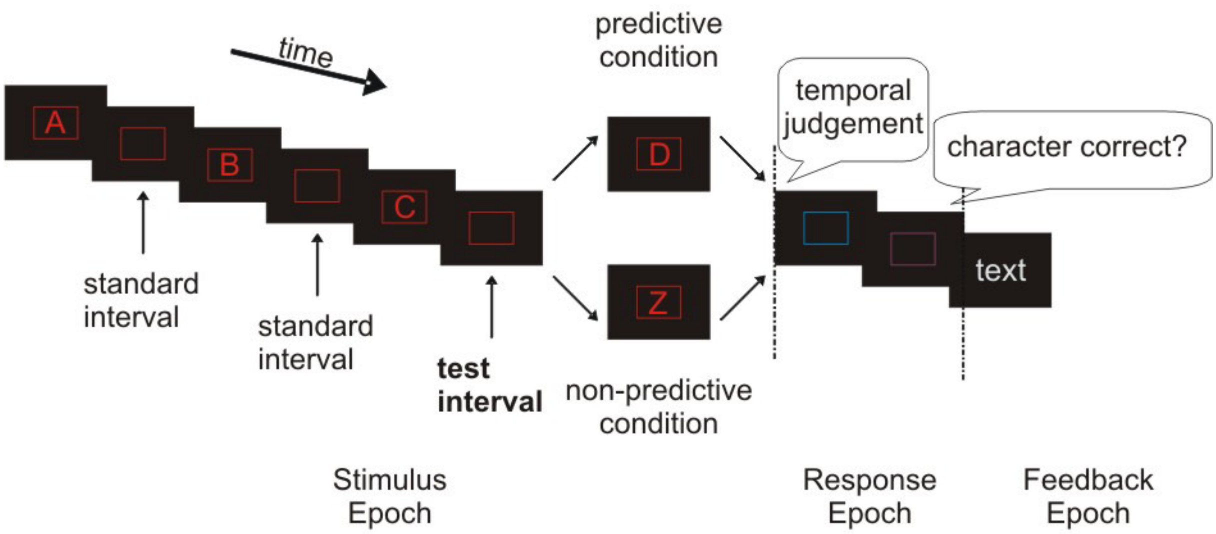

Figure 18: Experiment task design (#02).

The figure is almost identical to Figure 11. A novelty in task design is the written feedback ("text") at the end of each trial. Feedback refers to the control task (alphabetical task). Other novelties apply to strategies, which are not visible in this illustration of the task design. Standard intervals are displayed for 1,000 ms and characters for 300 ms. Test interval varied for the pretest and was fixed for the main experiment. For all other details, see caption of Figure 6 (basic concept).

Design #02: Results and Implications

We conducted this experiment on 34 subjects. Incorrect answers to the control-task were immediately labeled invalid. Therefore, there was no need to exclude those trials after the experiment, as was necessary in the previous design. Prior to calculating the psychometric function, we excluded all trials with eye movements and blinks in a critical period, as described in the “eye tracking” section of “material and methods.” In addition, we applied the inclusion criteria for subjects, as described in further detail in the “experiment procedure” section of “material and methods.” In brief, those were:

- Complete coverage of psychometric function
- JND <400 ms in pretest[14]
- 100 or more valid trials left without blinks and saccades (for Design #02 eye movement data was analyzed post-hoc)
- Reliable fit of psychometric function[15]

After this process of exclusion, datasets of 10 subjects were available for analysis (n=10). The majority of excluded subjects were excluded due to eye movements. Other exclusion criteria also contributed, e.g. coverage of psychometric function.

Figure 19 exemplifies and illustrates the relevance of these exclusion criteria. The importance of covering the full psychometric function (0–100%) and at least the range of 25–75%, as is relevant for reliably estimating both PSE and JND, can be seen: The pre-PSE value of subject #01 was estimated as *minus* 57ms. This is misleading and not interpretable. Because the pre-PSE was fed into the main experiment as a central value, to measure a range of pre-PSE +/− 250 ms, a negative value is also inoperative. Negative values cannot be used as a test interval length. The last character would need to appear before the preceding character disappeared, which is impossible. If less than 25—75% of the psychometric function are covered, the PSE and JND derived from it come with high uncertainty. The

[14] JND was used as follows:
Old pretest (sub. 1–8): JND <200ms (proceed experiment and include data into data analysis)
New pretest (sub. 9 onward): JND < 150 ms: proceed experiment and include data into data analysis; JND 150 ms −400ms: proceed experiment and check, if data can reasonably be included in data analysis, e.g. by checking main experiment’s JND < 400 ms; JND > 400 ms, subject is not allowed to proceed with the experiment.

[15] P-value was corrected from $p<0.05$ to $p<0.0013$, which is a conservative correction. $P<0.005$ was the original value. Calculation as follows: 0.005 / (20*2) = 0.0013, where 20 was the number of participants, of whom two were chosen for a two-tailed testing.

values might be misleading, as is the case for the functions illustrated in Figure 19. In addition, both psychometric functions in Figure 19 appear to be close to linearity instead of sigmoidal. A typical sigmoidal psychometric function is shown in the "hypothetical results" section.

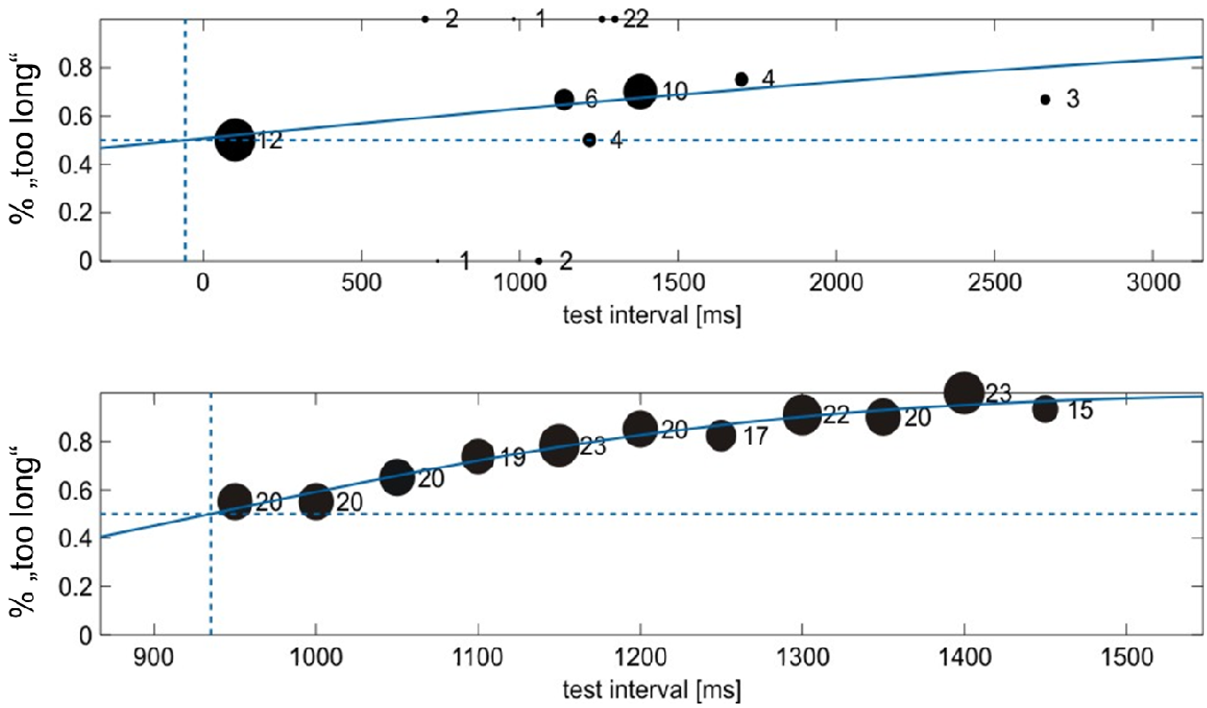

Figure 19: Psychometric functions of two subjects who did not cover the full psychometric function. Upper function from Sub #01 (pretest) covers the area from 50% to 100% proportion of temporal judgements judged as "too long." Lower function from Sub #17 (main experiment) covers 55% to 100% proportion of temporal judgements judged as "too long." PSE value (50% "too long") is at the intersection of the blue lines. Horizontal dashed line indicates that 50% judged as "too long." Vertical dashed line indicates the intersection of the function with the horizontal line. PSE values are −57 [ms] (upper) and 934 [ms] (lower).

In addition, JND (test interval at 75% − interval at 25% proportion-corrected value divided by two) is required to be below 400 ms in the pretest data. In the upper panel, the psychometric function of pretest data is shown. The respective JND value is 2,152 ms. This subject (#01) failed to pass any of the exclusion criteria of the pretest. The function in the lower panel (sub #17) reflects a JND value below 400 ms for the main experiment. This dataset was excluded because it did not cover the full psychometric function.

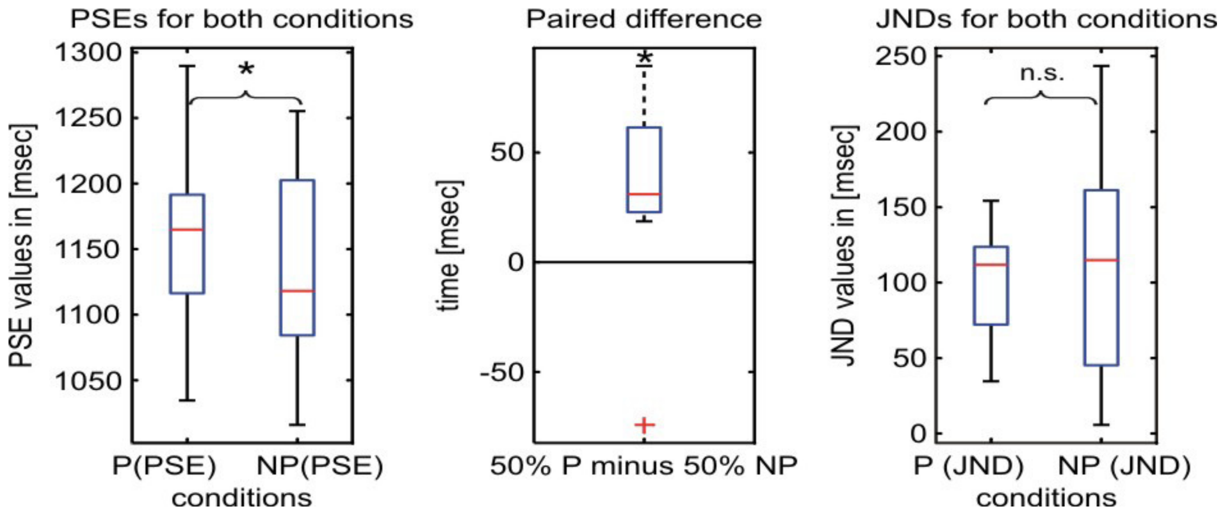

Figure 20: Data from paradigm #02.

The left graph gives PSEs of both conditions with significant difference (p= 0.043, two-tailed paired t-test, alpha = 0.05). Median values (red) are 1,165 ms (P) and 1,118 ms (NP). The hypothesized effect of a difference of conditions is given. The same interpretation can be taken from the central boxplot. Paired differences clearly cluster above zero, median value (red) is above zero at 31 ms (mean 33,3 ms). Asterisk indicates significance (same values and data as on the left). JNDs (right) show broad variance and differed not significantly (p=0.5, two-tailed paired t-test, alpha = 0.05). Therefore, higher values for P-PSE cannot be explained by altered JND values. (n=10; All boxplots show quartiles in blue and median in red. Whiskers give 1.5 times the interquartile range; outliers are displayed with a red + sign. Asterisk indicates significance. Only data sets that passed the exclusion criteria were analyzed.)

The hypothesized effect of the difference between predictive and non-predictive conditions was present in the data. This was a positive result. In detail, PSE values of both conditions showed a statistical difference (see Figure 20) (boxplot, left) (p= 0.043, two-tailed paired t-test, alpha = 0.05). The paired difference amounted to 33,3 ms (median), which was just within the expected range. Neuronal processing delays for vision take around 80 ms to 100 ms (see “introduction”). We assumed a predictive effect on the perceptual level to contribute to partial delay compensation, but not to cover the full delay. Therefore, the value of 33 ms matched this expectation. The paired difference (Figure 20, center) is another way to visualize the difference between P-PSE and NP-PSE. In addition, in a combined scatterplot with boxplots (see Figure 21), the PSE values of each of the 10 subjects can be seen. The diagonal line indicates the PSE values at which P and NP condition equal out—or, in other words, where the paired difference is zero. Nine out of 10 data sets are clearly below this line, indicating positive paired differences. This is echoed in Figure 20 (center), with a boxplot above zero (median +31ms). Just one data point is above the diagonal line in Figure 21. This outlier can be located in Figure 20 (center) as the red cross at minus 74 ms. JNDs show large values, but due to exclusion criteria, none of them are above 400 ms. They all cluster around 100 ms, with the highest value close to 250 ms. The JNDs of the predictive and non-predictive conditions were not significantly different (Figure 20, right). Therefore, higher values for predictive PSEs cannot be explained by altered JND values (Wirxel & Lindner, 2012). As in the previous design, PSE data reveal a tendency to settle at values above 1,000 ms.

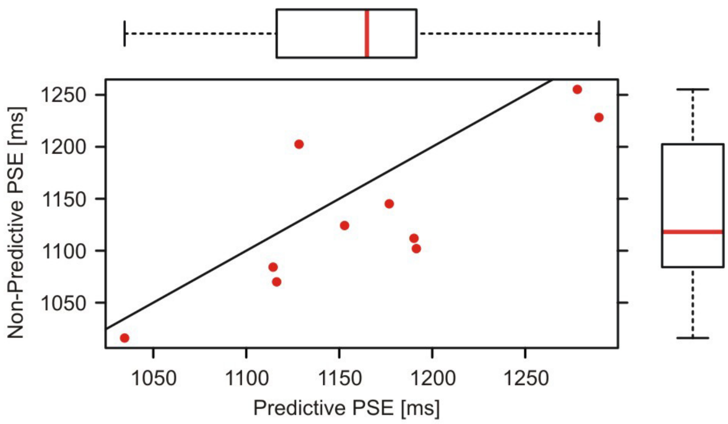

Figure 21: PSEs from predictive and non-predictive conditions (Design #02, n=10).
The diagonal line indicates the point at which PSE values for both conditions (P, NP) equal out. Every dot below the diagonal represents a subject whose predictive PSE was longer than the non-predictive PSE. Nine out of ten subjects showed larger PSE values in P conditions (data below line). There was a significant shift of the PSE toward larger values for predictable sequences (median shift: 33 ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Median values (central bars) are 1,165 ms (P) and 1,118 ms (NP).

Design #02 has two disadvantages. The first is a high dropout rate of subjects. Only 10 out of 34 subjects were left for analysis. This makes data acquisition ineffective and expensive. Therefore, a better design is required.

The second disadvantage might even be suitable to disqualify the whole paradigm, because one could argue the effect found in Paradigm #02 might be based on an artifact: Each character was displayed for 300 ms, including the last character, which was either predictive or non-predictive (see Figure 18). In the non-predictive case, the subject “knew” that the last character was the last character as soon as it appeared on screen (onset). This is because it did not follow the alphabetical order.

In the predictive case, the last character could not be identified as such at its onset. The end of the sequence was only obvious from the response screen's appearance. This led to a systematic difference between the two conditions. Because the last character is displayed for 300 ms, subjects are informed about the last character being the last character 300 ms earlier in the non-predictive condition, compared to predictive condition. Whether this systematic difference is suitable for producing any effect—especially the effect we found in the data—cannot be verified. Nonetheless, the 300ms issue requires an experiment re-design.

As a control for task difficulty, we tested for effects in reaction time and accuracy. By the term accuracy, we mean the percentage of correct answers to the control task (alphabetical task). By the term reaction time, we mean the time elapsed during the response phases, until the subject presses a response-button. If one condition was more challenging than the other, this might be reflected in reaction time and accuracy. If task difficulty increases, accuracy and/ or reaction time should decrease.
We did not find any difference in reaction time and/or accuracy between the two conditions. Accuracy was high across both conditions and across all sequence lengths. It sampled around 95% (P: 94–99%; NP: 95–99%, n=10, data not shown). Reaction times for temporal judgement task varied between 1,000 ms and 2,000 ms and reaction times for the control task ranged between 300 ms and 700 ms (n=10, data not shown). Accuracy neither increased nor decreased with increasing sequence length. With accuracies of about 95%, the lack of significant differences might be also due to a ceiling effect, as there might simply have been no space for further improvement.

We tested for an interaction of sequence length and condition (P; NP) based on PSE values (see Figure 22). There was no interaction (p=0.82). Length had no effect (p=0.65, Anova, n=10), whereas condition had an effect (p=0.035, Anova, n=10). We considered sequences of four and five characters as short while seven and eight characters were considered as long. We did not take sequences with six characters into account. Interestingly, there was a side effect. With increasing sequence length, PSE values decreased and converged to standard interval length (1,000 ms). The longer the sequence length, the more accurate was the temporal judgment.

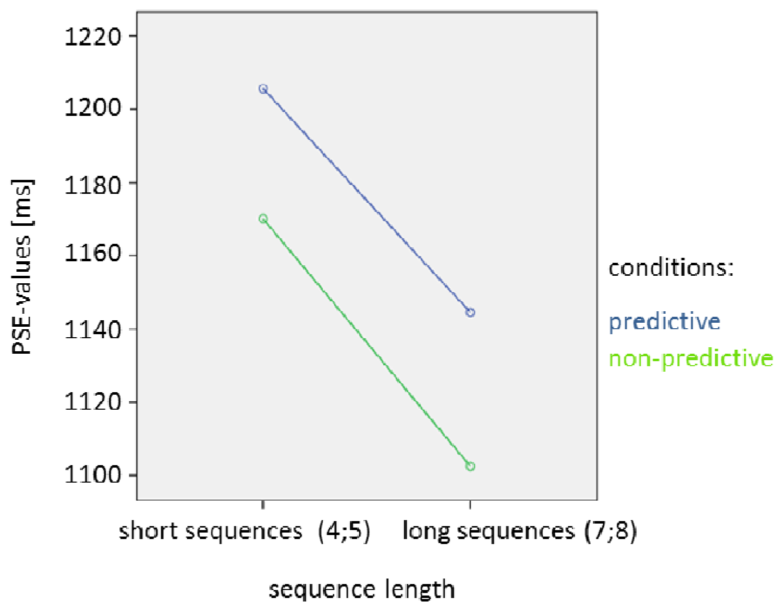

Figure 22: Interaction of sequence length and condition on PSE values in Design #02. Blue (P) and green (NP) lines are parallel, showing that the discrepancy between P-PSE vs NP-PSE does *not* change with sequence length. PSE values decrease with increasing sequence length. Thus, with increasing number of standard intervals, the temporal judgements get closer to standard interval length (1,000 ms).

Accurate in this case means conforming to standard interval length, and not conforming to smaller JNDs/ steeper slopes (which we would refer to as precision). This speaks in favor of a perfection of temporal judgment with increasing number of standard intervals. For absolute temporal judgments, subjects seemed to have a "learning curve." The longer the sequence, the more accurate/close to the standard interval length was the judgment. Perhaps judging the interval length was more difficult than answering the alphabetical control task; therefore, there was space for improvement.

We expected an interaction of length and condition. Our expectation was a relative increase of P PSE values as compared to NP PSE values with increasing sequence length/number of standard intervals. The rationale is that with increasing sequence length, subjects saw more standard intervals, which might be beneficial for prediction. If length had such an effect, this spoke in favor of a real predictive effect based on contrasting our experiment conditions, not in favor of a putative 300ms artifact. However, we failed to obtain such an effect (no interaction). This neither verifies nor falsifies the possibility that our positive result was caused by the 300ms issue. Hence, we need an approach to avoid the 300ms issue.

Furthermore, we expected the discrepancy of P-PSE vs NP-PSE to increase with sequence length. According to this expectation, blue line (P) and green (NP) line in Figure 22 should shift apart with increasing sequence length. As a matter of fact, there was no such effect, as can be seen in Figure 22. Blue (P) and green (NP) lines are parallel.

Still, a promising aspect of this design is the fact that with the novel pretest, we improved the selection of adequate test interval lengths (based on the pretest PSE). Still we had to exclude subjects, because also this pretest did not consider all possible “shortcomings,” as is illustrated in Figure 19 (see upper psychometric function of subject #01). In the remaining subjects, however, the data recorded in the main experiment clustered much more precisely within the relevant range of the psychometric function. According to the spontaneous feedback of the subjects after the experiment, they also felt encouraged by the immediate feedback on their performance. Because of the 300ms issue and the high dropout rate, an adjustment was required. The adjustment aimed to reduce the dropouts and to address the 300ms issue.

Design #03—First Control

Design #03: Concept

The substantial novelty of Design #03 was to address the 300ms issue. Hence, we added a visual cue to identify the last character in both conditions equally. Therefore, we replaced the red rectangle by a red circle, surrounding only the last character and in both conditions (see Figure 23). This label informed the subjects that the last character would be the last character at its onset. The 300ms difference in registering that this was the last character ceased to exist between the two conditions. An additional novelty was the invention of “online” eye-tracking, including immediate feedback on eye movement and blink performance at the end of each trial. A third novelty was the display of a fixation cross during standard and test intervals. Online eye-tracking with feedback and the fixation cross aimed to reduce the vast dropout rate due to insufficient eye-movement and blink performance. A second text displayed at the end of the feedback epoch provided feedback for eye movements (Figure 23). Feedback was either “*gute Augenbewegung*” (good eye movements) or “*schlechte Augenbewegung*” (bad eye movements). As a fourth novelty, circles substituted rectangles during the response phase. Colors remained unchanged. As before, standard intervals were displayed for 1,000 ms and characters for 300 ms.

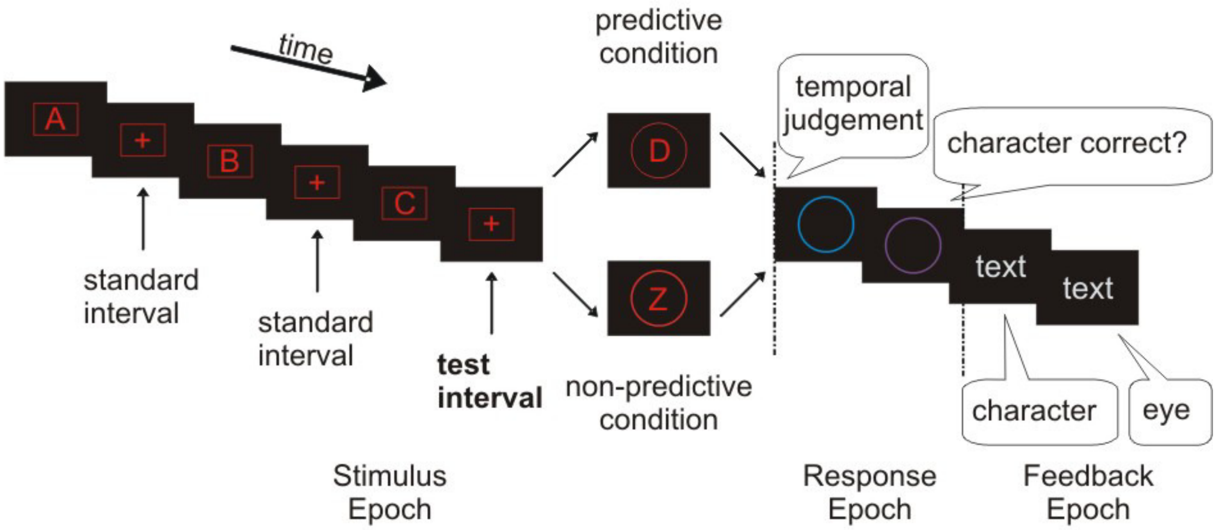

Figure 23: Experiment task design (#03).

Four changes are visible in our task design as compared to Design #02. First, the last character for both conditions is identified by a surrounding circle instead of a rectangle. Second, written feedback on eye movement performance (for simplicity labeled as "text" in the figure) was provided at the end of the feedback epoch. Third, a fixation cross was displayed during standard and test intervals. Fourth, circles substituted rectangles during response phase. For all other details, see caption of Figure 6 (basic concept).

Design #03: Results and Implications

We conducted the experiment six times to get data. Due to the novel online analysis of eye movements, there was no need to process eye data before applying the probit analysis for fitting the psychometric function on valid trials. Trials including eye movements and blinks during the period, which we defined to be critical because the test became invalid immediately (see "eye tracking section" and "exclusion and inclusion criteria"). We measured only six subjects in this design. The reasons are detailed below. Two subjects dropped out during the pretest. The remaining four subjects showed a broad variation in their results. Since the overall data set is too little for statistical analysis, we just plotted the available data and provided the according data set (see Figure 24 and Table 1). Paired differences are −17.7 ms, 35.6 ms, −10.6, −26.6 (median is −14ms). Those values give the distance from the diagonal line in Figure 24. Positive and negative values define positions on either side of the line. Positive values are below and negative values above the line.

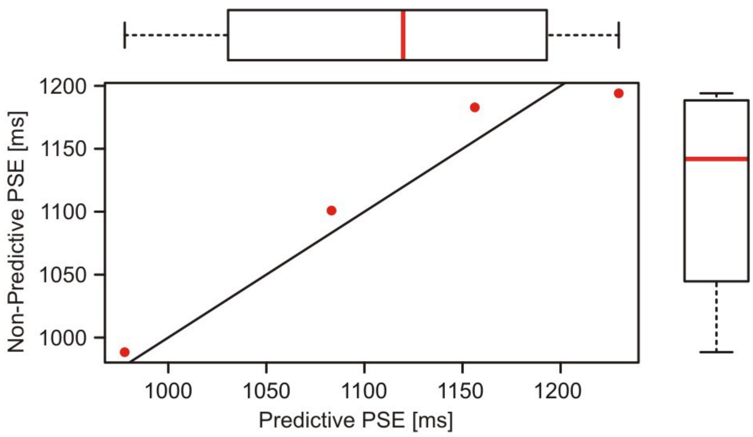

Figure 24: PSEs from predictive and non-predictive conditions (Design #03, n=4).
The diagonal line indicates the point at which PSE values for both conditions (P, NP) equal out. Every dot below the diagonal represents a subject whose predictive PSE was longer than the non-predictive PSE. One out of four subjects showed larger PSE values in P conditions (data below line). The data set was too small to allow a statistical analysis. There was no shift of the PSE toward larger values for predictable sequences (median shift: -14ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Median values (central bars) are 1,120 ms (P) and 1,142ms (NP).

For transparency, we have listed the individual data of all four subjects in Table 1.

Time 3

	Paired Difference P PSE - NP PSE	50% P	50% NP	JND P	JND NP
Sub3	-17.6928	1083.213	1100.906	67.1351	44.4385
Sub4	35.5685	1229.634	1194.065	141.8832	176.1347
Sub5	-10.5976	977.7883	988.3859	155.8678	143.5911
Sub6	-26.6089	1156.269	1182.878	92.2672	73.8626
Median	-14.1452	1119.741	1141.892	117.0752	108.7269

Table 1: Data from Design #03.
Dataset was too small for statistical analysis. Therefore, raw data was shown.

Although we planned #03 as a full experiment and not as pilot, we stopped data acquisition after six subjects due to the broad variation, including negative PSE values. The possible reasons for this variation are considered below. In any case, according to our hypothesis, we had expected positive differences in PSE values. Dropout rate was a secondary problem. Three out of four subjects showed a tendency in the opposite direction of the effect found in Design #02. In case those subjects were just outliers, which is possible, and additional subjects would show positive PSE values, we would have to collect more data than planned. With the predefined number of n=10 subjects and the significance level of α=0.05 it was impossible to get a significant result. Acquiring more data with this design would still have led to a null result. Since this is simply a result that does not support our main hypothesis, and is also not appropriate to reject the hypothesis, it is of no consequence. Thus, further data acquisition would consume time and resources without the chance to provide insights into the addressed hypothesis. This design is possibly inappropriate to address the hypothesis.

Reasons for the absence of the effect, found in Design #02 need to be discussed. One possible explanation is that the effect seen in Design #02 was an artifact caused by the 300ms issue. The 300ms issue was addressed by the new design. Hence, it would be of no surprise not to find an artificial effect again in Design #03 (if it was truly artificial). To verify or to falsify the real existence of the hypothesized effect, as seen in #02, a continuation of data collection in Design #03 would have been of no use. A second possible explanation for the absence of the effect is a change in predictability. With the unique circular label of the last character, this character got a framing that was different from all the preceding characters, standards, and test intervals (compare Figure 18 [Design #02] and Figure 23 [Design #03]). This change in framing could be interpreted as a change in predictability. When taken Gestalt-psychology into account, the whole set of symbols has to be taken into account, not only the characters. Each frame, whether it is a rectangle or a circle, surrounds the character and is part of the set of symbols. Accordingly, changing the shape of a frame from rectangular to circular would change the predictability of the whole set of symbols (compare Table 2). For the purpose of discrimination, predictability of frames is indicated in lower case (p; np). Predictability of last characters (conditions) is written in capital letters.

frame / condition	p (#01; #02)	np (#03)
P	P & p	**P & np**
NP	NP & p	NP & np

Table 2: Predictability of condition and last frame for Designs #01–03.
Conditions are written in columns (P, NP). Lower case letters indicate the predictability of the last frame/ label (p; np). Predictability of last frame is written in rows (np= circle; p=rectangle). For frames, p refers to Designs #01 and #02, np refers to Design #03. Design #03 provides a P&np (bold) combination of elements, instead of a fully predictive P&p combination.

We did not intend this change in predictability when designing #03. In Designs #01 and #02, the predictive condition (P) consisted of a fully predictable set of symbols (compare Table 2). Solely the non-predictive (NP) condition did not, as was intended. In case Gestalt-psychology was right and the full set of symbols were taken into account by each subject, we would have changed the predictability of the whole trial sequence in an unintended way. If this was true, a possible explanation for the loss of the effect seen in #02 could be a kind of a "startle effect." A surprising symbol in a sequence of expected, predictable symbols might cause a startle effect and require more cognitive resources. It is possible that this hypothetical additional demand of cognitive resources is suitable to cover/mask or to abolish the temporal effect, as seen in Design #02. This is speculative, but may still explain the findings.

Working around the unintended change in predictability and addressing the 300ms issue at the same moment is a challenge.
Fixating the sequence length with a fixed number of characters, (e.g. four) could counteract. The subjects would know when the last character appears. This holds true for both conditions. Nevertheless, there are deficits as well. The subjects might find it too demanding to focus also on sequence length. Their task for the experiment is to judge the interval length, focus on the content of characters to answer the alphabetical control-task, and to avoid eye movements. Based on the voluntary report of subjects across all designs, this is already pretty demanding, although it may sound easy.
There is another deficit of a fixed sequence length. In the first place, we decided to keep the sequence length variable to make sure the subjects are not able to "foresee" the last character from the first character. For example, if sequence length were four, starting with "A" would indicate "D" to be the last character, while "M" indicates "P" to be the last character and so on. But subjects should not be able to initially estimate the last character. We wanted the sequence to succeed character by character. The subjects should be forced and motivated to focus on the content of each character. Subjects should include the characters' content to assemble a prediction character by character. This is what it is all about.
If the subjects made a temporal judgment based on (a fixed) sequence length or on rhythm, without taking characters' content into account, the experiment would no longer be adequate to measure predictability. The whole experiment would be

obsolete. In the surprising case of an effect showing the hypothesized characteristics, we would not know how to interpret it, because it would not be based on predictability or it could be based on a different kind of predictability. Thus, other approaches to work around this issue are required.

Design #04—Second Control

Design #04: Concept

The novelties of this design were the following: First, the removal of the frame during the stimulus epoch (see Figure 25). Second was a replacement of color-coded frames during the response epoch. The response epoch was indicated by written words instead. “*Zeit”* (time) reminded subjects to make a temporal judgement. “Alphabet” (alphabet) reminded subjects to judge whether or not the last character was correct according to the alphabetical order (control task). Third was an introduction of sound to announce the end of sequence at the onset of the penultimate character. Sound replay and character display started simultaneously. We chose a typical computer “click” sound. Headphones were provided and the subjects adjusted sound volume to a comfortable level before training started.

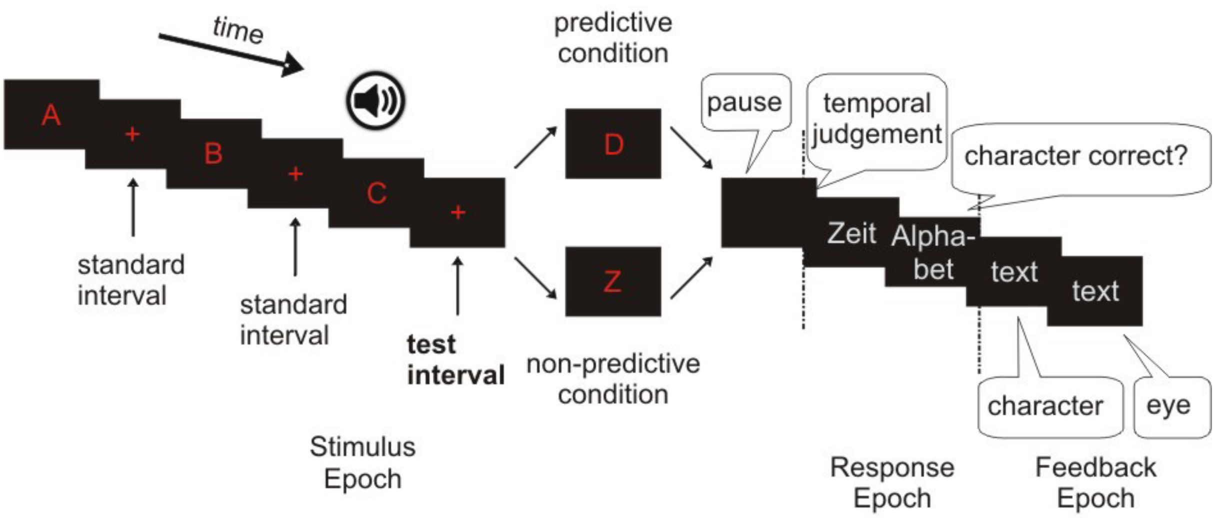

Figure 25: Experiment task design (#04).

Three changes in task design are visible as compared to Design #03. First, we removed all frames surrounding the characters. Second, response screens were labelled by written words. Third, a novel pause-interval (200 ms) was added after the last character and before the response epoch starts. A non-visible change is the addition of a click-sound, which occurred at the onset of the penultimate character. The sound icon symbolizes this sound. For all other details, see caption of Figure 6 (basic concept).

Fourth, we introduced a pause between the end of the sequence and the beginning of the response epoch (see Figure 25, the respective frame is labelled “pause”). The length of the novel pause-interval was 200 ms. (The very first training session, “training without eye tracking” had a 100ms pause. Subsequent “training with eye tracking,” “pretest,” and “main experiment” had pause-intervals of 200 ms. This was an unintended inconsistency, due to a typographical error. Since the very first training without eye tracking was only used to provide procedural knowledge to the subject, this should be of no consequence.) As before, standard intervals were displayed for 1000 ms and characters for 300 ms.
Replacing the frames surrounding characters during the stimulus epoch was an approach to prevent the predictive/non-predictive Gestalt problem of the previous Design #03 (see table 2). For this reason, we removed all frames during both the stimulus epoch as well as the response epoch. In addition, it seemed to be much easier for subjects to get the response instruction in written words instead of color-coded symbols. However, the change in the response epoch from color-coded frames to written words led to another concern: Written words consist of characters, as does the stimulus sequence. Thus, to clearly separate characters from the stimulus-sequence and characters from the response epoch, the pause-interval was introduced.

Introducing sound was a way to address the 300-ms issue. As a reminder, in Design #02, the non-predictive last character could be identified as such already 300 ms earlier than the last character for the predictive case. Labelling the last character for both conditions in the same way, as for Design #03, had unintended side effects and therefore did not solve this issue. Sound was another approach to “label” the end of sequence. On the one hand, there was the necessity for an identical label for both conditions to create a systematic difference between both conditions. On the other hand, labelling should not result in a P/np-combination of characters and labels (which was the case for Design #03, compare Table 2). Labelling the penultimate character in the same way for both conditions seemed a viable option. The subject would know when the sequence would end and the *last* characters of both conditions were expected <u>not</u> to contain P/np-combinations of character and label. Instead, the P/np-combination of character and label would occur at the penultimate character and—in the same way—across both conditions. Therefore, the penultimate letter got labelled by sound, whereas the last character was not labelled.

Design #04: Results and Implications

We conducted the experiment with 12 subjects. A total of 10 subjects were included according to our criteria (n=10). One subject was excluded because of a high JND-value (see exclusion criteria). Another subject withdrew his/her consent after the pretest. The remaining 10 subjects showed no significant effect (p=0.28, two-tailed paired t-test, alpha = 0.05). PSE values showed a broad variation from below 600 ms to above 1,000 ms (see Figure 26).

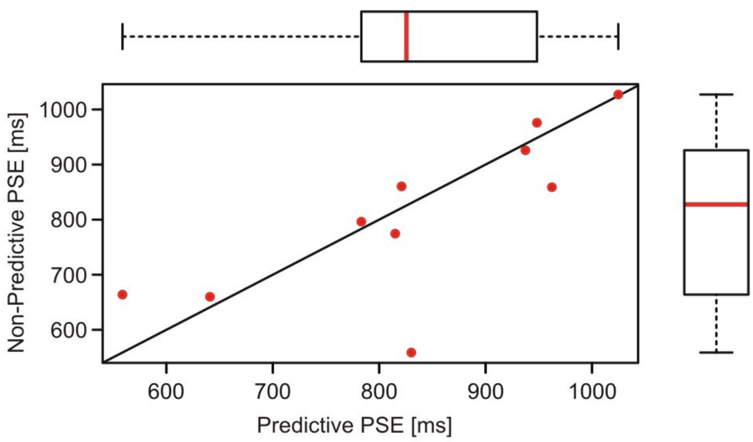

Figure 26: PSEs from predictive and non-predictive conditions (Design #04, n=10).
The diagonal line indicates the point at which PSE values for both conditions (P, NP) equal out. Every dot below the diagonal represents a subject whose predictive PSE was longer than the non-predictive PSE. Four out of 10 subjects showed larger PSE values in P conditions (data below line). There was no significant shift of the PSE toward larger values for predictable sequences (median shift: -8ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Mean values (central bars) are 832 ms (P) and 810 ms (NP).

The null result has at least two potential and speculative, explanations. The first explanation assumes that there was no incentive for forming predictions. It builds on the fact that the last character was clearly identified, and whether this character could be predicted or not was basically chance level (50%). In other words, from the moment when a subject listened to the click-sound, chances for the next character to be predictive or non-predictive were equally likely, irrespective of sequence length. Hence, it would have been sufficient for the subjects to form a prediction only for the last character (if at all). Moreover, this prediction would fail in 50% of the cases. Thus, this design did not provide any incentive for forming predictions. Instead, in previous designs, the subjects did not know beforehand when the last character would appear and thus supposedly formed predictions for each character—predictions that were confirmed in the vast majority of cases and thus positively reinforced.

One might argue that subjects felt no need to predict, which therefore explains the null result in the current design (#04). However, this train of thoughts is based on the idea that one can *solely* expect a benefit from prediction, if the prediction is likely to be correct. One might further argue that with a higher share of predictive trials, this design might have worked out. The second potential explanation for the null result is speculative as well. It assumes that one particular combination of the last character and "label" might have been disadvantageous for forming a prediction. As a reminder, capital letters (P, NP) indicate the predictability of the conditions (last character), while lower case letters indicate the predictability of the last character's label (p; np). The argumentation is as follows: the last characters of both conditions should <u>not</u> contain P/np-combinations of character and label (here, the presence and absence of sound), as mentioned before in the context of Design #03. The absence of sound at the last character's onset was there for both conditions. We expected this equality to solve the issue. Possible combinations for the last character thus were a.) predictive character (P) without sound and therefore without any label (P/--) and b.) non-predictive character (NP) also without sound and therefore without any label (NP/--). Whether this lack of a label (sound) itself is perceived as predictable (p) or non-predictable (np) is not clear. Therefore, both cases can be described as P/p and NP/p or as P/np and P/np (see table 3). If we were assuming that the lack of sound was interpreted as non-predictive, this led to a combination of P/np (Table 3, lower table). Such a combination of P/np has been speculated to be disadvantageous for forming a prediction in the context of Design #03 (table 2) already.

label / condition	p
P	P & p
NP	NP & p

label / condition	np
P	**P & np**
NP	NP & np

Table 3: Predictability of condition and "label" for Design #04.
Conditions are written in columns (P, NP). Lower case letters in rows indicate the predictability of the label (p; np). Upper table: assumed absence of sound (last character) is perceived as predictive (p). Lower table: assumed absence of sound (last character) is perceived as non-predictive (np). The lower case represents a combination of a predictive character and a non-predictive label (bold). Combination is the same as in Table 2 (Design #03).

Despite the null result, the general decrease in PSE values, as compared to the previous designs, was a remarkable finding. All previous designs were purely visual and PSE values ranged clearly above 1,000 ms. Given a standard interval length of 1,000 ms, PSE values ranging from 1,100 ms to 1,200 ms were surprisingly high in our purely "visual" designs. For the audio-visual design (#04), PSE values dropped for the first time below 1,000 ms. Median values were 832 ms (P) and 810 ms (NP). To verify whether this decrease was caused by the auditory modality, further investigation would be required. This finding is likely to be caused by the auditory/visual cross-modal design, as it was the major novelty.

Each of the explanations is speculative. To verify or falsify them, subsequent experiments would be required. Explaining the null result or the drop of PSE values below standard-interval length is not an integral part of my main research questions; hence, we do not get into that matter. Instead, to reproduce the positive result seen in Design #02 already, another adjustment was required. The auditory sound implemented in Design #04 was most likely to be the source of the null result and the side effect. The adjustment led to Design #05, which was again a purely visual design.

Design #05—Third Control

Design #05: Concept

The novelties of this design were the following: First, we removed the click-sound. Second, the sequence length was fixed to six characters. Due to this, subjects "knew" when the sequence ended. Hence, a special cue for the last character was not required anymore (see Figure 27). Please note, as in the previous Design #04, in the training session without eye tracking, the pause-interval lasted 100 ms only. During data acquisition, the pause-interval was 200 ms. As before, standard intervals were displayed for 1000 ms and characters for 300 ms.

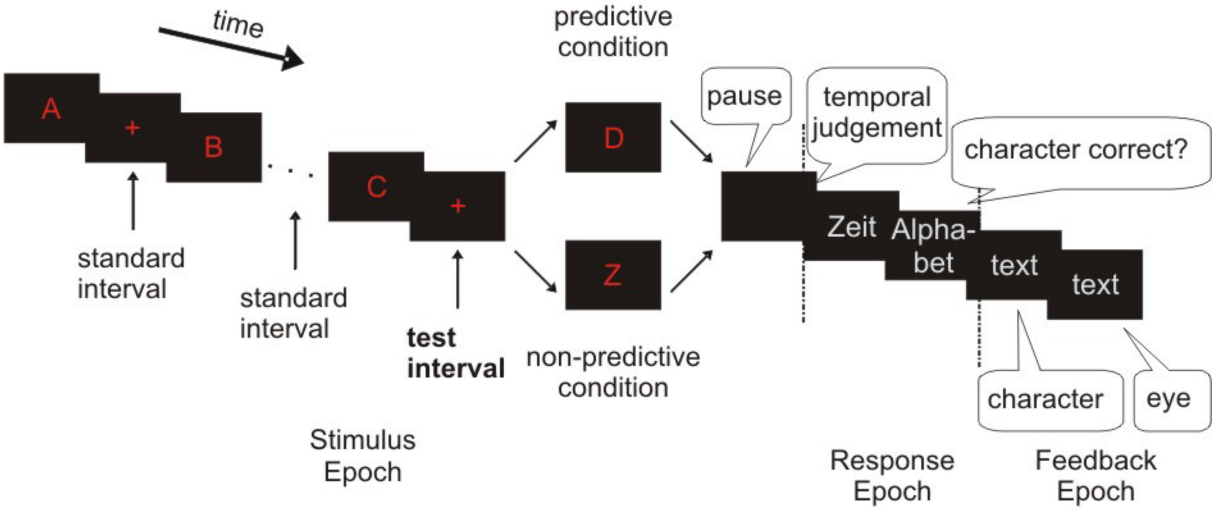

Figure 27: Experiment task design (#05).

A visible change in task design is the fixed sequence length of six characters, symbolized by the dotted line. No sound symbol is shown, because no sound was replayed in this design anymore. The pause interval remained, as was introduced for Design #04. For all other details see caption of Figure 6 (basic concept).

Design #05: Results and Implications

We conducted the experiment in nine subjects. Two subjects quit after pretest, seven subjects remained for analysis (n=7). The remaining seven subjects showed no significant effect (p=0.81, two-tailed paired t-test, alpha = 0.05). PSE values showed a broad variation from below 1,032 ms to 1,705 ms (see Figure 28). The paired difference between predictive and non-predictive PSEs was −38ms (median) with four negative and three positive values. Individual paired differences ranged from −50 ms to +186 ms. A positive value reflects a larger PSE value for predictive condition than for non-predictive one. A positive shift is what we expected according to our hypothesis. Unfortunately, a majority of subjects in this design exhibited negative values. Interestingly, PSE values cluster clearly above 1,000 ms, again ranging between 1,100 ms and 1,200 ms. This further supports the idea that the decrease of PSE in Design #04 was caused by the auditory cue, signaling the upcoming end of the sequence.

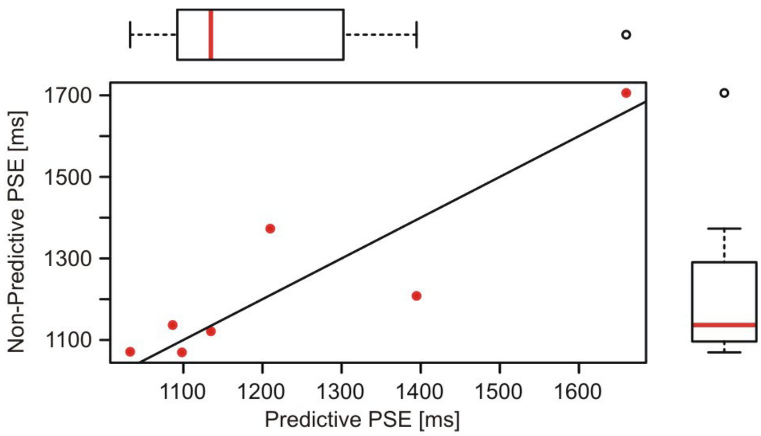

Figure 28: PSEs from predictive and non-predictive conditions (Design #05, n=7).
The diagonal line indicates the point at which PSE values for both conditions (P, NP) equal out. Every dot below the diagonal represents a subject whose predictive PSE was longer than the non-predictive PSE. Three out of seven subjects showed larger PSE values in P conditions (data below line). There was no significant shift of the PSE toward larger values for predictable sequences (median shift: −38ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Median values (central bars) are 1,135 ms (P) and 1,137 ms (NP). Empty black circles mark the outliers.

All other designs were purely visual and provided PSE values above 1,000 ms. The cross-modal design was the only one with PSE values clustering below 1,000 ms.
By using a fixed sequence length, we avoided any kind of extra-labelling for the last or penultimate character. Hence, we had no more possible combinations of P&np, as was the case for previous designs #03 and #04.
The 300ms issue remains debatable. We assumed that this issue could potentially lead to false-positive results (false-positive significant larger PSE values for predictive condition). With the fixed sequence length, we intended to address this issue. First of all, Design #05 revealed a negative result. Hence, it is needless to discuss the 300ms difference in the context of Design #05. Although the 300ms difference in design was still existing, a positive result was missing. Without such a result, there is no need to distinguish between false-positive result and correct positive result. Taken together, whether fixating sequence length is an adequate approach to the 300ms issue cannot be evaluated from these data, because of the negative result.
Although it seems intuitive that a longer sequence length is beneficial to create a prediction, data from Design #02 do *not* support this intuition (see section “Design #02: Results and Implications”). Interaction analysis (Design #02) reveals no

interaction of sequence length and condition (P; NP) based on PSE values (see Figure 22). There was no interaction (p=0.82). Length had no effect (p=0.65, Anova, n=10). Therefore, there was no necessity for Design #05 to choose the maximal sequence length (eight was the maximum number of characters in previous designs). Sequence length was fixed to six characters (Design #05).

Similar to the preceding Design #04, there is a 50% chance level for P/ NP conditions in the present design. If one assumes that subjects were able to recall/count six characters in this particular design, the chances for each condition at the sixth character were 50%. Subjects knew that straight from the beginning. As pointed out for Design #04, one might argue that subjects felt no need to predict—they were right in 50% of the cases. This train of thoughts is based on the idea that one can only expect a benefit from prediction, if the prediction is likely to be correct. Only then do subjects focus on predictable information. For the recent design (#05), subjects did benefit from prediction in just 50% of the cases. One might argue that with a higher share of predictive trials this design might have worked out.

Design #06—Ultimate Design (IR and MEG setup)

We used two different setups for Design #06: First, the “infrared lab” (IR) psychophysics setup, as for the previous designs, and second, an MEG setup, additionally allowing the simultaneous recording of brain activity. Due to physical restrictions of the MEG environment, technical modifications of the respective psychophysical setup were inevitable.

Design #06: Concept for IR and MEG Design

This design was strictly based on Design #02 (see Figure 18 and Figure 29) (#06-IR). We had a positive result in #02, which we would like to reproduce without the 300ms issue. The 300ms issue was the motivation to keep changing Designs #03–06. To address the 300ms issue, we went for a novel approach. We stopped our efforts to announce the end of sequence for both conditions in the same way because of all the aforementioned concerns. Instead, we decided to display each character very briefly, for one frame only. With such a brief display time, we expected to reduce the 300ms issue to the length of one display frame, while assuming that an increase of the beneficial effect of prediction due to a higher detection threshold would be even larger.

To reiterate, this approach did not eliminate the 300ms issue completely. Instead, it meant a reduction of the confounding variable from 300 ms to close to nothing: Depending on the technical limitation of the setup, one frame lasted 10 ms at IR lab setup (100 Hz) and 16.67 ms at MEG setup (60 Hz). Hence the systematic difference between conditions was reduced to 10 ms (IR) and 16.67ms (MEG) respectively. The effect we expected, based on data from Design #02, ranged from 30 ms to 35 ms. A systematical difference of 10 ms (IR) or 16.67ms (MEG) should hardly cause an artificial "predictive" effect of twice the size.

Again, the concern relating to the 300 ms issue was a temporal advantage for the NP condition, because the subjects could realize at NP characters' *onset* that it was the end of sequence. Instead, subjects had to wait in the predictive condition until the response screen was displayed. This temporal difference was 300 ms. Reducing it to 10 ms (IR) or 16.67ms (MEG) and recovering an effect of even greater size (i.e. the PSE-shift is larger than one frame) would speak in favor of an effect based on condition (NP vs. P) and therefore would be based on predictability.

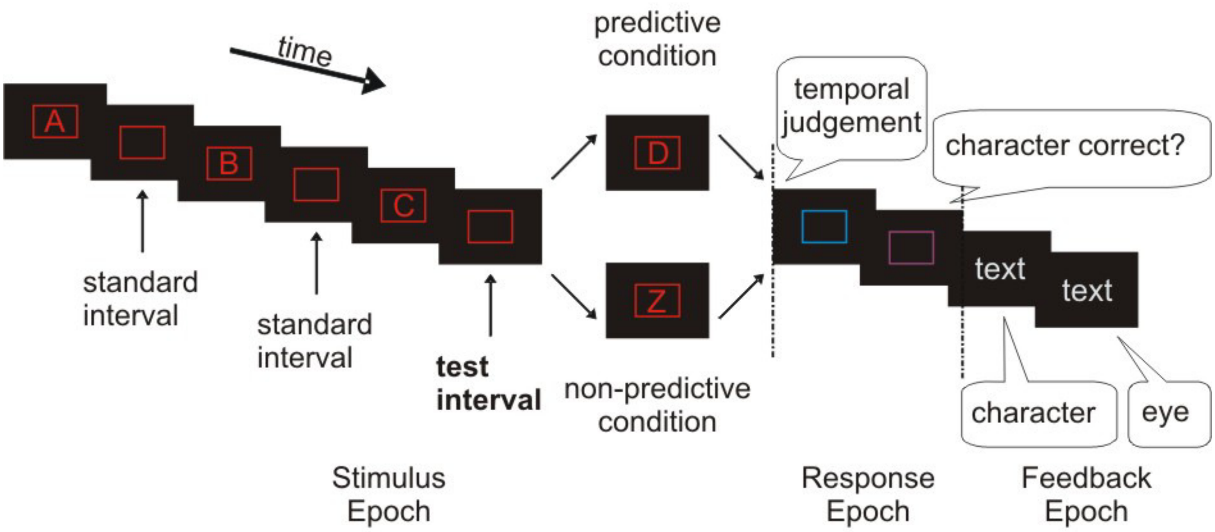

Figure 29: Experiment task design (#06 IR-setup).
Changes in task design: sequence length varied from four to eight characters as in Designs #01–04. Character display length was reduced to one display frame (10 ms @ 100Hz). The pause interval was removed as in Designs #01–04 (pause interval was only implemented in previous Design #05). For all other details, see caption of Figure 6 (basic concept).

Design #06: Results and Implications of Measurements at IR setup

We conducted the experiment on 10 subjects. One subject quit due to backpain occurring after the pretest. Another subject could not avoid making eye movements during the pretest. Therefore, we decided to end the experiment on this subject after the pretest, as the main experiment would have taken far too long to yield a sufficient number of trials due to our online eye-movement control. The remaining eight subjects (n=8) showed a significant effect (p=0.04, one-tailed paired t-test, alpha = 0.05). This is a positive result; the effect from Design #02 was reproduced. PSE values varied from 900 ms up to 1,303 ms (see Figure 30).

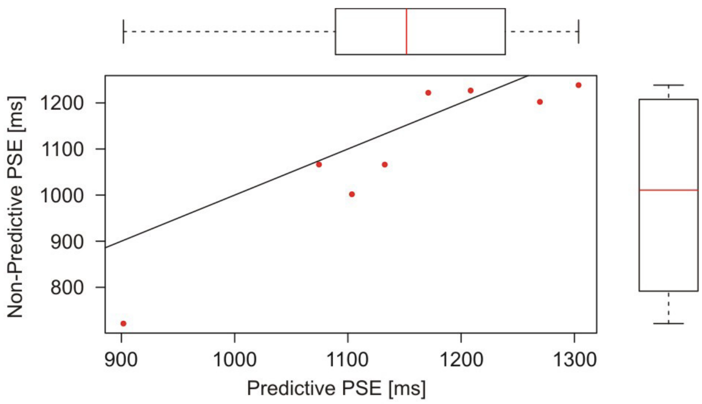

Figure 30: PSEs from predictive and non-predictive conditions (Design #06-IR, n=8).
The diagonal line indicates the point at which PSE values for the two conditions (P, NP) equal out. Every dot below the diagonal represents a subject who perceived the predictive PSE longer than the non-predictive PSE. Six out of eight subjects showed larger PSE values in P conditions (data below line). There was a significant shift of the PSE toward larger values for predictable sequences (median shift: −66ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Median values (central bars) are 1,151 ms (P) and 1,134 ms (NP).

The previously seen effect of Design #02 recurred. With the current design, the effect even increased. The median PSE shift was 66ms and therefore twice as high as in Design #02 (median shift 31 ms). Data in #06 were not normally distributed; therefore, we provide median instead of mean values. JND values of predictive and non-predictive conditions were not significantly different (p=0.45, t-test, two-tailed, paired). Hence, difference in PSE values cannot be explained by altered JNDs. With respect to PSE values, an effect of 60–70 ms is within the hypothesized range (also see general discussion), which was estimated on the basis of knowledge about processing delays in the visual system (~100 ms). In addition, we assumed that

prediction might not fully compensate for this delay at a perceptual level, but may contribute partially to its compensation (see introduction).
The results of this experiment were highly promising and in support of our hypothesis. Therefore, we continued establishing an MEG-compatible version of this design. We aimed to study whether this difference in temporal perception of predictive vs. non-predictive events is also reflected in electrophysiological data. Specifically, we wanted to check for relative differences between conditions in VEPs while focusing both on differences in (a) latencies as well as (b) amplitudes. Both would be suitable to further support our hypothesis.

As a control for task difficulty, we tested for effects in reaction time and accuracy (n=8 subjects). By the term accuracy, we mean the percentage of correct answers for the control task (alphabetical task). If one condition was more challenging than the other, this might be reflected in reaction time and accuracy. If task difficulty increases, accuracy should decrease and reaction time should increase.
Reaction time for temporal judgement task samples around 640–820 ms (P: 670–820 ms, NP: 640–760 ms). Reaction times were pretty similar across conditions, with slightly higher values in predictive conditions. Interestingly, reaction times for the same task were much higher in Design #02 (1,000–2,000 ms).
Reaction time for control task ranged from 990 ms to 1,100 ms (P: 990–1,100 ms; NP: 990–1,050 ms). Interestingly, in Design #02, reaction times for control task were mainly between 300 ms and 700 ms. Taking into account that none of the experiments were designed as a reaction time task, as subjects were not instructed to respond as fast as possible but to be accurate, reaction time results should not be overestimated.
Accuracy was high across both conditions and across all sequence lengths. It ranged between 95% and 98% for P and between 89% and 98% for NP. The lowest value, 89%, belongs to a sequence length of five characters. All other sequence lengths, including the shortest sequence with four characters gives accuracy values of 96%-98%. Therefore the 89% value is most likely an outlier. Accuracy neither increased nor decreased with sequence length. With about 95% accuracy, this lack of an influence of sequence length might also be due to a ceiling effect.
If we were lucky, we could also answer the question regarding whether the relative shift between conditions was forward or backward in time. Although it was not part of

our hypothesis to detail the direction of the shift, we might be able to answer it. Specifically, is the shift of the PSE caused by the predictive condition being perceived earlier, or by the non-predictive condition being perceived later (because a prediction was violated), or by both effects? A point of reference would be needed to address this question, which is required to be neutral—i.e. it is not allowed to contain any predictive expectation or violation of a prediction. It needs to be neutral in terms of prediction and violation of prediction.

Design #06: Conceptual details of MEG setup

Next, we transferred Design #06 from the IR to the MEG setup. Changes were required due to technical constrains of the MEG environment. Most importantly, we wanted to keep the presentation period for each character at one frame (as short as possible). For visual stimulation in the MEG setup, we used a back-projection system via beamer instead of a CRT screen. This beamer did not support a 100Hz frame rate; it supported only 60Hz frame rate. Therefore, the presentation time of characters now was 16.67 ms. The length of the standard intervals remained unchanged (1,000 ms). As explained in the “experiment procedure” section, pre-PSE from the pretest was fed as central value into the main experiment. Due to temporal resolution of MEG setup the central value for the main experiment was rounded to multiples of 16.667 ms in whole numbers. Up to subject no. 7, central values were rounded to 0; 17; 33; 50; 83, and 100 ms. For subject no. 8 and the following subjects, central value was rounded to 50 ms (0; 50; 100).
Additionally, we introduced a baseline period before the first character appeared. We did this because for the MEG setting it was planned to provide physiological data, which could be used as point of reference or “baseline.” We chose a baseline length of 2000 ms while assuming that the first half or two third of this interval would not provide adequate data for the calculation of a point of reference or “baseline,” due to contamination from the previous trial, while still providing a sufficient time epoch for a baseline calculation.

To provide accurate stimulus-locked time-stamps as point of reference for MEG analysis, we introduced two markers, which were registered within the ACQ-data set from the MEG equipment. As markers, we used projected white squares on the stimulus screen that were locked to the stimulus sequence. These white squares were projected onto optical fibers that were connected to two photodiodes. The respective voltage data were recorded together with the MEG signals and stored in the ACQ files. As these markers were projected on the alloy frame of the back-projection screen, they were not visible to the subjects. The first marker appeared for one frame before the baseline period started. The second marker appeared together with the last character of the sequence (see white squares in Figure 27). Implementing these marks was important for data analysis, specifically the estimation of the VEPs.

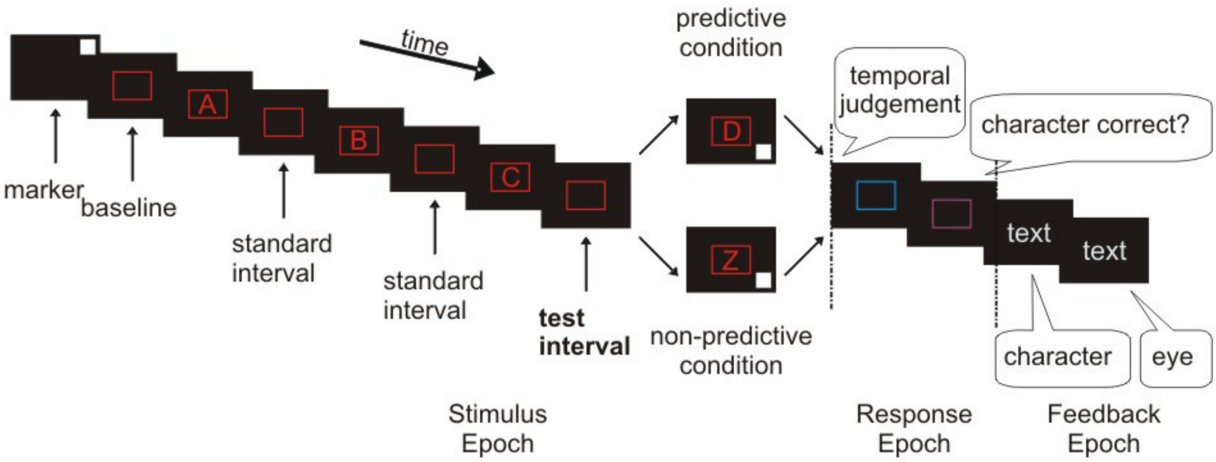

Figure 31: Experiment task Design #06 (MEG setup).
A change in design, which was not visible to the subject, was the introduction of temporal markers (white squares). The presentation time of characters was 16.67 ms (one frame). Standard intervals remained unchanged at 1,000 ms. For all other details, see caption of Figure 6 (basic concept).

Design #06: Results and Implications of Measurement at MEG setup

For the MEG setup, we split the experiment into 15-minute blocks and continued until 200 valid trials were completed. This was due to the huge amount of data (ACQ files). We conducted the experiment on 13 subjects. Two participants did not continue after the pretest. Therefore, 11 participants (n=11) ultimately performed the main experiment. For these 11 analyzed datasets, the PSE was 20 ms (median) higher for the predictive condition (median 1,139 ms) than for the non-predictive condition (median 1,117ms). This difference was marginally significant (p=0.06, one-tailed paired t-test). Remember that the original version of this design (#06-IR) revealed a clear significant effect.

The lack of significance (MEG version) might be caused by small differences between the two setups. The main differences were a lower frame rate of display and the introduction of a baseline interval of 2,000 ms (#06-MEG) instead of only 1,000 ms (#06-IR). Although we instructed subjects to judge the test interval length compared to the standard intervals, the longer baseline interval may have interfered with this judgment and caused the smaller effect size. Baseline was now twice as long as ISIs, whereas in the previous IR-design (#06-IR), the two were of the same length. If subjects had taken the baseline into account, prediction might have not developed properly. This might have caused the weakening of the effect. However, this is speculative.

As a control for task difficulty, we tested for effects in reaction time and accuracy (MEG; n=11 subjects). By the term accuracy, we mean the percentage of correct answers to the control task (alphabetical task). If one condition was more challenging than the other, this might be reflected in both reaction time and accuracy. If the task difficulty increases, accuracy should decrease and reaction time should increase.
Reaction times for the *temporal judgement task* ranged between 1,200 ms and 1,400 ms (P: 1,200–1,300 ms; NP: 1,330–1,360 ms). This range is in accordance with the reaction times of the same task in Design #02 (1000–2000 ms).
Reaction times for *control task* were around 700 ms (P: 690–750 ms; NP: 700–790 ms). Interestingly, when compared to the same task in both Design #06 (IR; 990–1100 ms) and Design #02 (300–700 ms), values are rather in accordance with Design #02.
Accuracy was high across both conditions and for all sequence lengths. It ranged around 97% (P: 95–98%, NP: 97–99%). As in Designs #06(IR) and #02, the accuracy neither increased nor decreased with increasing sequence length. With about 97% accuracy, this might be due to a ceiling effect: There was no space for further improvement. Within the tested designs (#02, #06-IR, #06-MEG), accuracy was high across both conditions and across all sequence lengths. Tested designs seemed to be equally manageable for participants.

To provide an overall picture of the results, we combined the behavioral data of temporal judgement task obtained in the MEG setup (#06-MEG) with those from the IR setup (#06-IR). Combined data are displayed in Figure 32.To allow the identification of data from the respective setups, we plotted data in different colors for each setup (see caption). Combined data revealed a significant difference between predictive PSEs (median 1,140 ms) and non-predictive PSEs (median 1,118 ms), when applying a one-sided paired t-test (p=0.04). The hypothesized effect again surfaced. With respect to JNDs, there was no significant difference in JNDs between conditions (p=0.2, t-test) for combined data. Hence, higher values for predictive PSEs cannot be explained by altered JND values.

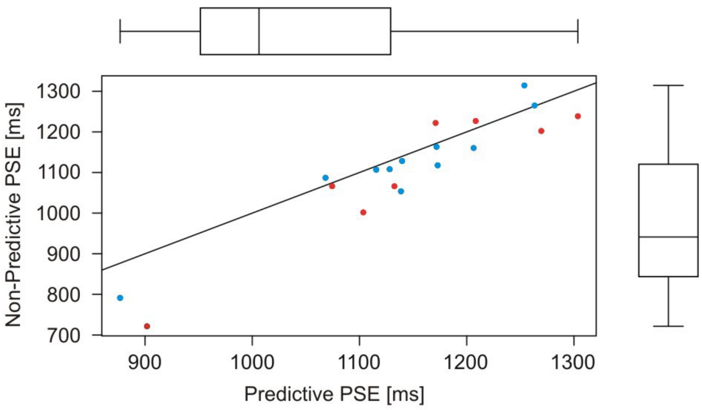

Figure 32: PSEs from predictive and non-predictive conditions from combined data (Design #06 IR and MEG, n=19).

Red dots: IR data, 100 Hz, n=8. Blue dots: MEG data, 60 Hz, n=11. Diagonal line indicates the point at which PSE values for the two conditions (P, NP) are equal. Every dot below the diagonal represents a subject who perceived the predictive PSE as longer than the non-predictive PSE. A total of 14 out of 19 subjects showed larger PSE values in P conditions (data below diagonal line). Overall, there was a significant shift of the PSEs toward larger values for predictable sequences (median shift: 20 ms). Boxplots capture the range between the 25th (Q1) and the 75th (Q3) percentile, and whiskers extend to extreme values. Median values (central bars) are 1,139 ms (P) and 1,117 ms (NP).

Design #06: MEG data

While the data of the MEG measurement (Design #06-MEG) was collected by Axel Lindner and myself, analysis of these data was part of a lab rotation. Therefore, analysis of MEG (ACQ) data was performed by a master's student (Katrina Quinn, 01.09.2014–07.11.2014). The reported results are built on her laboratory report. All other data—including behavioral data from the MEG study—were analyzed by me.

The intention of the MEG study was to investigate the relative differences between predictive and non-predictive VEPs. In particular, we were interested in the differences in time shifts of VEPs as well as differences in amplitudes.

MEG data preprocessing.

"Data analysis was performed using Fieldtrip (REF; http://www.ru.nl/neuroimaging/fieldtrip) running on MATLAB v.12b (The MathWorks). Each block of data was individually cut into trials before being baseline-corrected with respect to the whole trial, and high-pass filtered at 1Hz. Next, an independent component analysis (ICA) was run on the concatenated data and artifact components were identified visually based on their topography and time course. These components were subsequently removed from the data. The third step involved visual rejection of whole trials that appeared to deviate significantly from the average level of brain activity. In the fourth and final stage of preprocessing, the trials were cut into epochs of interest and the invalid trials were removed. The epochs were defined as 500 ms before stimulus onset to 1000 ms after the stimulus onset, and were cut separately for the first, second, penultimate, and last letters, and subsequently baseline-corrected individually (baseline interval: 200 ms prestimulus to stimulus onset)" (K. Quinn, lab report p. 9). Please see Figure 31 for the details of experiment task design.

MEG data analysis.

"All data analysis was carried out at the sensor level. For each participant, time-locked analyses were computed for each epoch of interest across all channels. This produced time-locked averages of the event-related potentials for 1) all trials, 2) predicted letter trials (P trials), and 3) [non-predicted] […] letter trials [NP trials] […]. These averages were then used to calculate the time-locked grand averages across all participants" (K. Quinn, lab report p. 9).

Results from MEG measurements.

Two main components were identified in both the first and the last character. The first component had its onset at ~130 ms (Figure 33). It was proposed to represent an early visual-evoked component (VEP), i.e. the M100 (magnetic form of the N100), elicited due to the presence of a visual stimulus. The second component had its onset at ~240 ms (Figure 33). This was proposed to reflect two different events for the first and the last character. For the first character, the second component (at ~240 ms) was proposed to represent a late VEP (e.g. the P200). For the last character, the second component (~240 ms) was proposed to represent the color change of the rectangle from blue to purple during the response phase (paraphrase of K. Quinn, lab report, p.10).

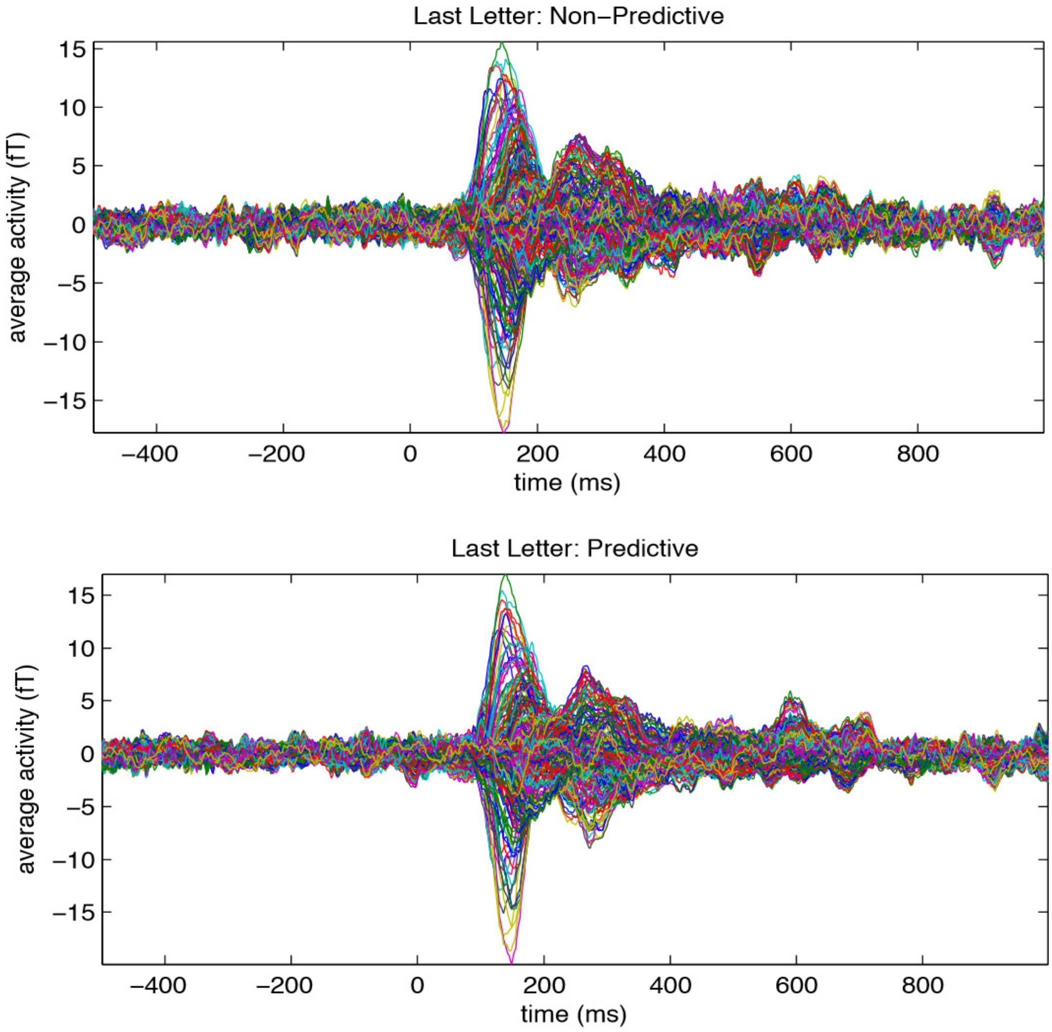

Figure 33: "Grand averaged waveforms for the last letter, plotted for each individual channel. Results are shown for non-predictive (top panel) and predictive (bottom panel) conditions." (Figure and caption taken from K. Quinn, lab report, with permission.)

The root mean squared (RMS) evoked response potentials (ERPs) for the first and last letters for both conditions (P and NP) are illustrated in Figure 34 and respective activity topographies are given in Figure 35.

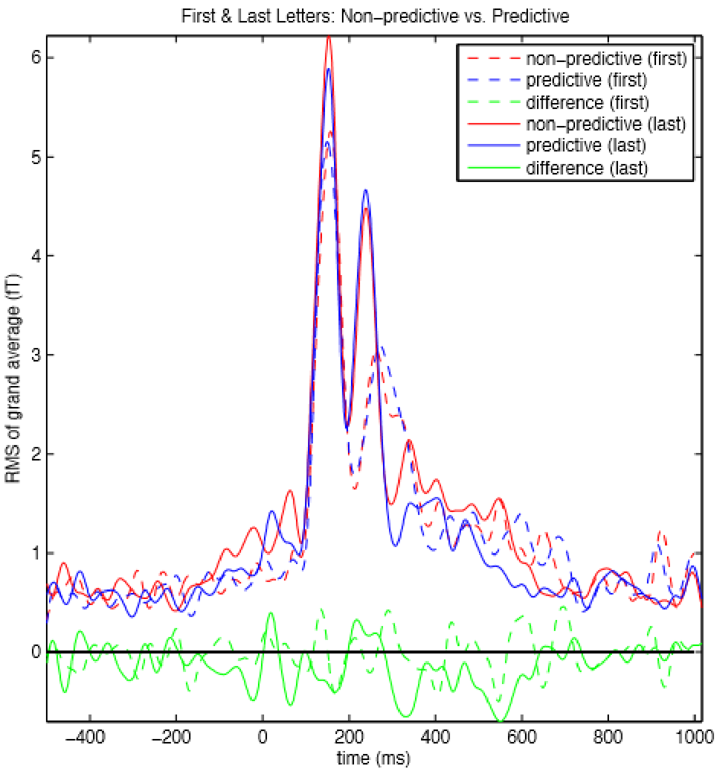

Figure 34: Grand averaged RMS ERPS (across all channels) for the first and last letter, with predictive and non-predictive conditions overlaid."
(Figure and caption taken from K. Quinn, lab report, with permission.) Dashed lines illustrate first letters, continuous lines illustrate last letters.

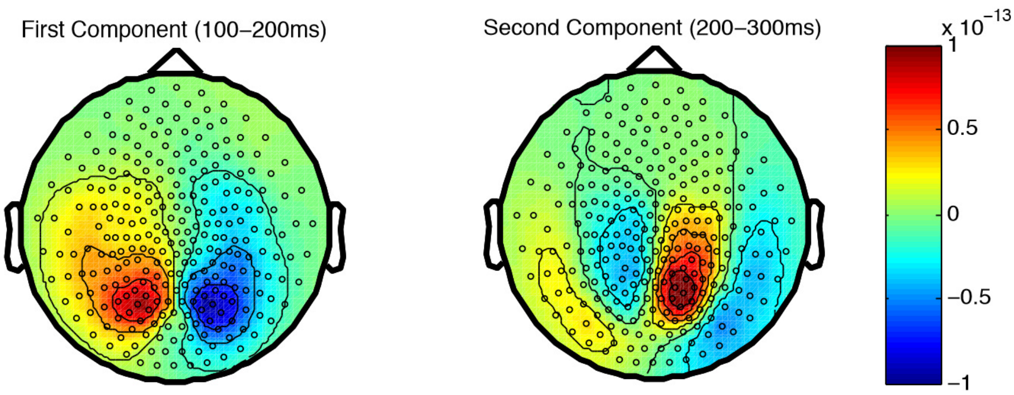

Figure 35: "Topographic maps of grand averaged ERPs collapsed across predictability conditions for the last letter.
On the left is the map for the first component (100–200 ms), and on the right is that for the second component (200–300 ms). Maps for the first letter are not shown, as they did not appear different to those from the last." Tesla is the unit used for the heatmap. (Figure and caption taken from K. Quinn, lab report, with permission.)

Relative differences between P and NP VEPs

Since we were interested in investigating the relative differences between predictive and non-predictive VEPs, analyses of ERPs of the last letter across both conditions were carried out. For the peaks of the first component (last letter), no significant effect was revealed in either timing (p=0.1669, paired sampled t-test) or amplitude (p=0.0646, paired sampled t-test). For the second component (last letter), no such analysis was carried out (paraphrase of K. Quinn, lab report). This was due to a “confounding effect of the response screen on the second component of the last letter, the timing of the second components for the first and last letters could not be compared” (K. Quinn, p. 13).

Time-shifts

Analyses that were thought to be more sensitive to small temporal differences were performed. These analyses focused on time shifts between predicted early VEPs and non-predicted early VEPs, including “1) averaging across occipital and parietal channels only, 2) topographic plotting of predictive vs. non-predictive latencies, 3) ICA of ERP components” (K. Quinn, lab report p.13). No significant difference in the timing of early VEPs between stimulus conditions (last character) was revealed through these analyses.

Potential mismatch negativity.

One interesting finding was a difference in activity between conditions for the last character, which potentially reflects the visual mismatch negativity (MMN). MMN is an electrophysiological correlate of the detection of *unpredicted* and unattended changes in our environment (Stefanics, Kremlácek, & Czigler, 2014). It is a component of event-related potentials (ERP), associated with the presentation of an odd stimulus in an otherwise regular/steady sequence of stimuli. MMN can be investigated by the use of either EEG or MEG. The MEG equivalent of MMN is termed as MMNm (Näätänen, Paavilainen, Rinne, & Alho, 2007). Intensive research on MMN in auditory modality has been performed (Stefanics et al., 2014), mostly by applying EEG.

For visual modality, there is much less literature available, since visual MMN (vMMN) is a comparatively younger research topic. Visual MMN (vMMN) is a (negative) ERP component that peaks around 200–400 ms after the onset of a visual stimulus (Kimura et al., 2011). Kimura and colleagues assess vMMN as an “evidence for the existence of unintentional prediction about the next state of a visual object in the immediate future on the basis of its temporal context," which they term "unintentional temporal-context-based prediction in vision." For Kimura et al., the prediction-error account is the most plausible approach to vMMN (not a memory-mismatch), as illustrated in Figure 36. According to this prediction-error account, a vMMN reflects prediction-error responses when the present event is incongruent with events that are anticipated on the basis of sequential regularities. These regularities are embedded in the temporal structure or context of the visual object. Multiple events are required to extract regularities and form a predictive model (see Figure 36). Hence, a single event cannot lead to vMMN. A similar approach also exists for auditory MMN.

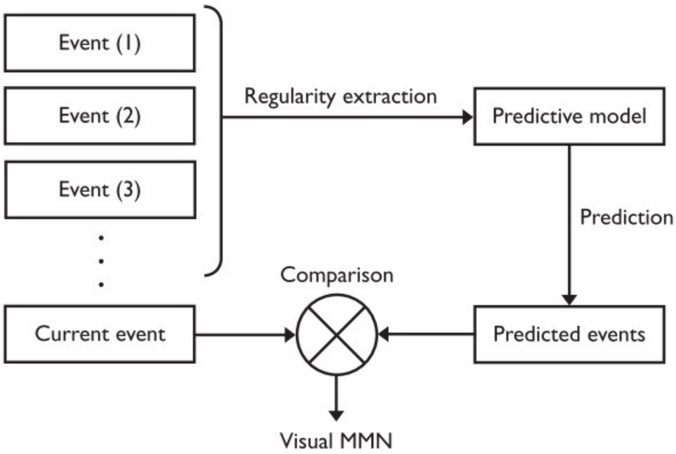

Figure 36: Schematic illustration of the prediction-error account (vMMN).
"A model for explaining the elicitation of visual mismatch negativity (vMMN) from the perspective of prediction-error account. Sequential regularities embedded in the temporal context of a visual object are extracted. Predictive models encoding sequential regularities are established. Predictions about upcoming events are formed on the basis of predictive models. A current event is then compared with predicted events. When a current event is incongruent with predicted events, visual MMN is elicited. These successive processes can operate without the observer's intention." Figure and caption are taken from Kimura et al., 2011[16].

As mentioned before, an interesting finding of our study is a difference in activity between conditions for the last character, which potentially reflects vMMN. The first character did not reveal such a difference. For the last character, the difference was identified between 400–600 ms in the parietal region (see Figure 37 and respective topographic map in Figure 38), with a *specifically greater activity in the non-predictive* (M=28.369fT, SD=3.98fT) *vs. predictive* (M=24.627fT, SD=4.659fT) conditions. The mean root mean squared (RMS) activity of the mentioned time window was calculated for both conditions, separately for the first and last letters. RMS implies that all activities are made positive. Averaging activity can lead to canceling out of the activity, because measurements from different gradiometers are combined. Based on

[16] With permission from Wolters Kluwer Health, Inc., license number 4153191293171.

the aforementioned calculation, a paired-sample t-test was performed and revealed a *significant difference between the unpredicted and predicted conditions* (t=4.3608, df=10, p=0.001). For the first character, no significant difference was found (t=−0.6077, df=10, p=0.5569). Therefore this component was suggested to represent a potential MMN component (paraphrase of K. Quinn, lab report, p. 13). According to Kimura et al. (2011), the vMMN peaks around 200–400 ms after the onset of a visual stimulus, whereas in our study the difference in activity between conditions for the last character peaks between 400–600 ms. This suggests that either we did not detect vMMN or we detected vMMN in an atypical time frame. The latter case would suggest that vMMN does not have a fixed time frame of occurrence. We considered the peak to be a *potential* vMMN.

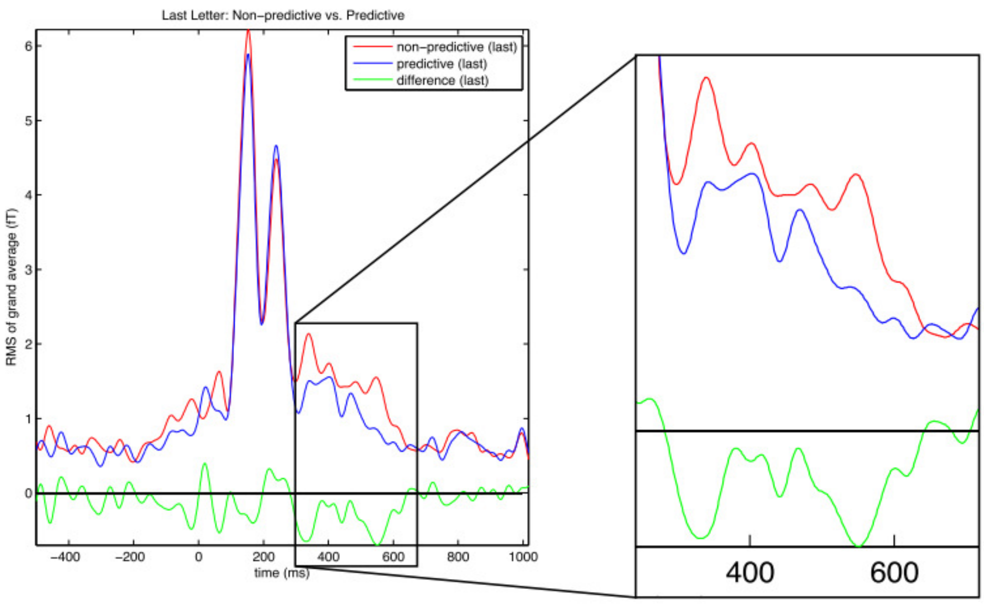

Figure 37: RMS evoked response potentials for the last letter (highlighted between 400 and 600 ms). Colors: red: non-predictive (last character), blue: predictive (last character), green: difference. RMS-ERPs are higher in the non-predictive condition for the mentioned time window. (Figure taken from K. Quinn, lab report, with permission.)

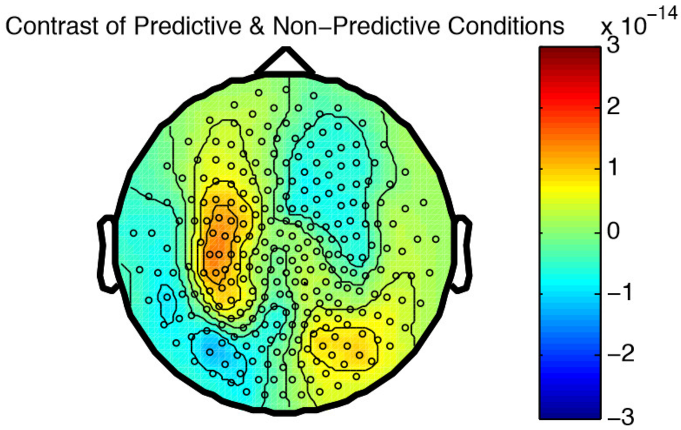

Figure 38: Topographic map of the predictive versus non-predictive contrasts (between 400 and 600 ms). Tesla is the unit used for the heatmap.

Eventually, “we were interested in the effects of the predictable stimulus stream on ERPs, independent of whether the last stimulus was predicted/[non]-predicted […]” (K. Quinn lab report). Therefore, trials were collapsed across conditions. We investigated the effects of predictable stimulus streams on ERPs, irrespective of whether the last visual stimulus was predictive or not (see Figure 39). For this data analysis, RMS ERPs of first, second, penultimate, and last characters were compared. For each successive character in sequence there was an increase in the first component’s amplitude (see Figure 40). When applying statistics, this trend did not prove to be significant ($p=0.1442$, one-tailed t-test). In detail, data for this analysis were treated as follows: “[W]e first normalized the data to values between 0 and 1 for each participant, collapsed the data across participants (11 participants x 4 letters = 44 values), and correlated the peak amplitude of the first component with letter position ($r= 0.2238$). The reason for our collapsing the data in this way was to provide as many values for the correlation as possible” (K. Quinn, lab report, p. 14–15).

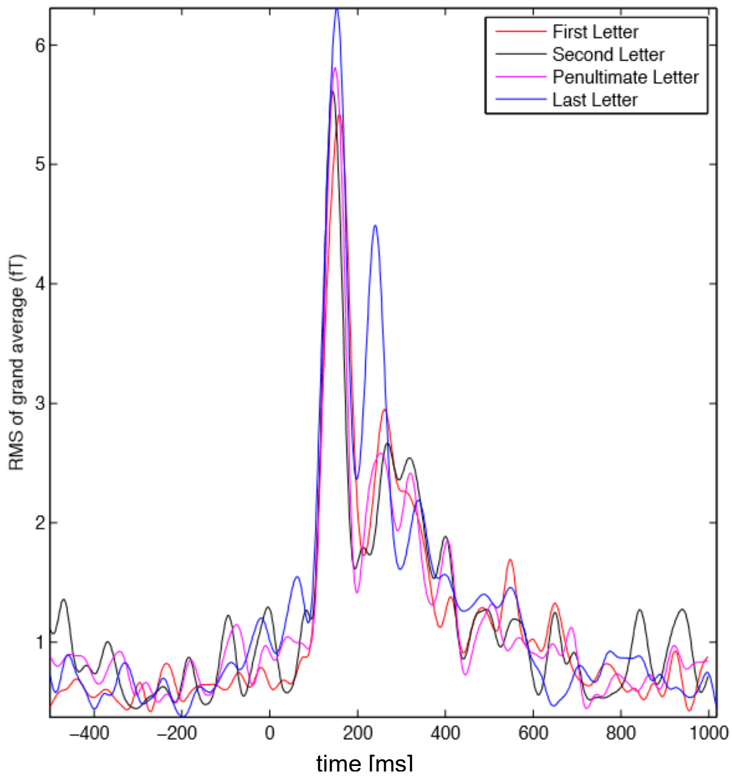

Figure 39: “RMS ERPs for first, second, penultimate, and last letters, averaged across predictability condition.” (Figure and caption taken from K. Quinn, lab report, with permission.)

Interestingly, with each successive character in a sequence, we found the trend of increasing amplitude in its respective VEP (first component), even though this trend was not significant (see Figure 40). It possibly reflects an increase in the facilitation of early visual processing (e.g. through prediction), because stimuli occur in sequence. For more details, see “general discussion.”

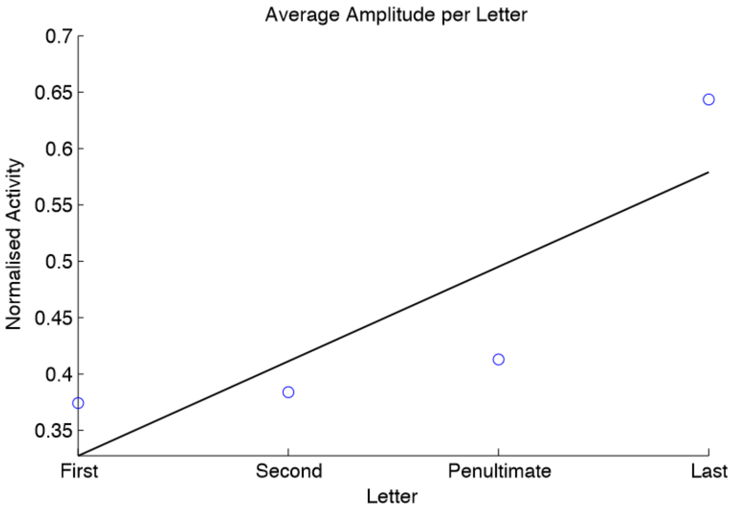

Figure 40: Correlation of letter position and amplitude of MRS ERPs. (Figure taken from K. Quinn, lab report, with permission.)

General Discussion

Findings on time perception across designs

Temporal compensation of neuronal processing delays may take place at the motor level (motor anticipation) or the perceptual level (see “introduction”). Prediction may be investigated by measurement at either the motor level or the perceptual level. Measurements at purely perceptual level are rare. To our knowledge, there was no compelling evidence for delay compensation at a strictly perceptual level. Therefore, we designed an experiment in which we focused exclusively on the perceptual level. Hence, we asked whether the predictability of a visual stimulus affects the time of perceived stimulus onset. In detail, we hypothesized that predictable visual stimuli have an earlier perceived onset compared to non-predictable stimuli. At the behavioral level, our results revealed a significant shift of PSEs toward larger values for predictable sequences. This shift is proof of a predictive effect. We expected subjects to perceive predictable characters earlier. This led to larger PSE values in *predictable* conditions than in *non-predictable* ones. As a consequence, the psychometric function for predictable stimuli was shifted to larger values (see Figure 7, “hypothetic results” section, and Wirxel & Lindner, 2012). This shift toward larger PSE values (P) was reproducible by two designs (#02 and #06). For the “original” design (#02), the mean shift was around 33 ms toward larger PSE values (significant, p=0.043). Differences in JNDs of predictable and non-predictable conditions were not significant. Therefore, the higher values for predictive PSEs could not be explained by altered JND values.

Design #06 (MEG and IR) revealed the same effect as we hypothesized and observed in Design #02. Mean shift of combined data (#06 IR and MEG) was 66 ms, and therefore about twice as high as in the “original” Design #02 (median shift 31 ms, mean shift 33 ms). JND values for the combined data-set again revealed no significant difference across conditions. Therefore, the higher values for predictive PSEs of the combined data set could not be explained by the altered JND values.

It should be noted that behavioral data from Design #06 measured at IR-lab setup revealed the hypothesized effect. Whereas data from Design #06 measured at MEG setup were only marginally significant. Combined data from both setups (#06-IR plus #06-MEG) again revealed the effect. Possibly, the IR-lab environment is more suitable to allow the detection of “fragile” psychophysical effects as compared to the MEG environment. This might be due to lightning conditions, comfort, higher temporal resolution of the stimulus display, etc. For an impression of the different environmental contexts for the subjects, see Figure 9 and Figure 10 (material and methods section).

Hence, we assume that the effect tends to be fragile and can easily be disrupted by perturbations. This can be taken from the fact that Design #06 revealed a significant hypothesized effect in one environment (IR), whereas the effect was only marginally significant ($p<0.1$) in the other environment (MEG). In addition, Designs #01, #03, #04, and #05 were also not suited to reveal the effect, although there might be also objective reasons for this (compare the specific discussions of each design in the previous section). The difference between Designs #02 and #06 referred mainly to the duration of the displayed visual stimuli. Characters were displayed for 300 ms (#02), 10 ms (#06-IR), and 16.67 ms (#06-MEG), respectively. The change from 300 ms to lower values like 10 ms or 16.6 ms was required due to a systematical difference between tested conditions. The non-predicted condition offered a temporal advantage: Subjects could realize by the onset of a non-predictive character that this was the end of the letter sequence, whereas in the predictive condition, subjects could only deduce that the sequence had ended by the onset of the response phase. Therefore, there is a temporal advantage for the non-predictive condition. Its magnitude is the duration of the last character’s display, which was 300 ms in Design #02, 10 ms in Design #06-IR, and 16.67 ms in Design #06-MEG. By shortening the display duration, we aimed to reduce the temporal advantage from 300 ms to 16.6 ms and 10 ms respectively. Still, there was a systematical advantage for the non-predictive condition, which became comparatively small. To check whether the systematical difference of one frame length might have caused the positive result, we did the following. For Design #06, we used the IR-dataset and subtracted the presentation duration of the last character (10 ms) from the paired difference (PSE P-PSE NP). Then, we checked if calculated values were different from zero. The result of this conservative testing was marginally significant (t-test, $p=0.07$, two-tailed, paired). We interpreted this as not confounding the significant shift in PSE values for the respective dataset.

The processing delay in vision is around 80–100 ms. We hypothesized not necessarily a full delay-compensation at the perceptual level. We expected perception to contribute to compensation. Hence, showing that predictable visual stimuli are in fact perceived earlier than non-predictable stimuli within the range of 31–66ms (depending on experiment setting) is a positive finding. The magnitude of 31–66ms seems plausible in the context of our study.

Another interesting finding on the behavioral level was that PSE values across conditions and designs ranged above the standard interval length, except for one design (#04). An unbiased observer would expect PSE values to range around the length of the standard ISI. Subjects were instructed to judge the length of the test ISI compared to the preceding standard ISIs. The length of standard ISI was always 1,000 ms. PSE values above 1,000 ms reflect an overestimation of test-interval length. So, why do subjects—irrespective of design and condition—overestimate the length of test intervals? According to Vierordt's law (1868), short intervals are overestimated and long intervals are underestimated. In between is the indifference interval, which is the interval length that is neither over- nor underestimated. Interestingly, across laboratories and even under fixed experiment conditions, there is not one single indifference interval that is valid for all subjects (Allan, 1979). One cannot assume one particular interval length to be overestimated in general. Hence, it was not necessarily expectable that ISIs of 1,000 ms length would be overestimated in our design(s). In addition, the original finding of Vierodt's law involved stimulus duration reproduction (Tse, Intriligator, Rivest, & Cavanagh, 2004), which is not the approach we choose for our study.
One might speculate that an interval length of 1,000 ms under experiment conditions of our study were in general overestimated in accordance with Vierordt's law. Solely in Design #04, PSE values across conditions ranged around 800 ms—much below the standard interval length of 1,000 ms. This could either be interpreted as a modulation of Vierordt's law by the experiment changes we implemented, like the auditory signal, or it could speak against an involvement of Vierordt's law.
Another—also speculative—account that could explain the general overestimation of test-interval length is the time-order error (TOE). The TOE refers to an influence in presentation order in a discrimination task. It was first described by Fechner (1860), who revealed a systematical asymmetry for comparative judgements when investigating JNDs in weight estimation. Subjects systematically overestimated the weight of the second stimulus, leading to an order-dependent bias in their JND

(Hairston & Nagarajan, 2007). In general, it is possible to over– or underestimate the first stimulus—the so-called negative and positive TOE (Hellström, 1985). One might speculate that TOE could have caused the overestimation of the test-interval. Again, it would be hard to consistently interpret our data by a TOE as PSE values in some contexts were ranging around 800 ms (Design #04), even though the standard interval length did not change.

Irrespective of the underlying mechanism for PSE values above and below standard ISI values— Vierordt's law, TOE, or another mechanism—it might have contributed to the difficulties with Design #01. This design was carried out without a pretest. Although an adaptive staircase procedure was applied to determine test-interval length, starting-points were chosen under the assumption that PSEs would range around the standard ISI value of 1,000 ms. This was obviously not the case. As an outcome of Design #01, we established a pretest based on an adaptive staircase procedure and fed the pretest PSE into the main experiment. The main experiment was carried out using an MCS. By doing so, we tailored the experiment to the perceptual range of each subject individually. We could do this because we were interested in the relative difference between predictive and non-predictive conditions but not in the question of whether PSE values were larger or smaller than standard ISIs.

At the behavioral level, reaction times were recorded. It should be noted that none of the experiments were designed to be speeded-reaction time tasks. We merely recorded and checked reaction times as a possible indicator of task difficulty. More importantly, our specific reaction time measure was not implemented to detect temporal benefits through predictive processing on the motor level. In fact, we did not aim to contribute to the corpus of work on delay compensation at the motor level. We aimed to contribute data supporting delay-compensation at a perceptual level. Therefore, we aimed to avoid any side effect from motoric measures on temporal judgement tasks. We used button presses; therefore, motor control was involved. But these button presses should not have interfered with the subjects' perception. Subjects made a comparative judgement on perceived interval length. The judgement itself was made without any motor involvement. When the perceptual decision was made, the subject only reported her/his decision via button presses. Because of this design, we are convinced that motoric actions, motor control, or any

motor prediction would not interfere with subjects' perceptual estimates. Nevertheless, we recorded reaction times, just to have them at hand—e.g. to check if they might reflect differences in task difficulty. Reaction times revealed no differences across conditions in Design #02. Taking into account the fact that accuracy of the control task was high for Design #02 as well as for both versions of Design #06, it is unlikely that ambiguous reaction times would reflect task difficulty. This is plausible when taking into account the fact that subjects were not instructed to make their judgments as fast as possible.
Because we controlled for eye movements, subjects were instructed not to blink during the stimulus epoch. Instead, they were explicitly informed that they were allowed to blink during the response phase. And they were allowed to use the full time of the two subsequent response phases of five seconds each) to full capacity. Therefore, at the onset of response phase, subjects decided how they were about to answer, but they might not have pressed the button immediately. Instead, they might have decided to relax their eyes for a moment, blinked to re-moisturize their eyes, and only then pressed the button(s). Therefore, any difference in reaction time might also reflect how dry subjects' eyes were and how well they used the response phase to relax their eyes for a moment. It should be noted, however, that high accuracy values that were indistinguishable between conditions render a systematic difference in task difficulty highly unlikely.

MEG results

With the MEG study, we aimed to shed light on the neural basis of the psychophysical effect we found in previous designs (#02 and #06 IR), which was the earlier subjective perception of predicted stimuli compared to unpredicted stimuli. We aimed to shed light on how this behavioral finding correlates with electrophysiology. In particular, we asked if predicted stimuli receive earlier sensory processing than unpredicted stimuli. Since we did not find such a difference in physiological (MEG) data, an alternative explanation is required to explain our behavioral results.

While we did not find a clear correlate of the prediction effect at the behavioral level, MEG data analysis still provides evidence for the formation of sensory predictions.
The evidence is given by the difference in activity between conditions for the last character, which is likely to reflect some kind of visual MMN arising from the

difference between a valid (P) vs. a violated (NP) prediction. The unpredicted condition thereby led to higher signal intensity. Accordingly, for the first character of a sequence, no difference was found, as no prediction may be formed in either condition. In previous studies, vMMN was found in the range of 200–400 ms (Kimura et al., 2011), whereas the difference in activity ranges from 400 ms to 600 ms in our study. Nevertheless, it might be a component of visual MMN. The presence of this potential vMMN component suggests—along with the high performance rates on answering the control task (alphabetical order task)—that subjects are able to make predictions on the perceived abstract (semantic) content of our stimuli.

For the auditory modality, the MMN is well-established (Näätänen, 1995), but it is much less so for the visual modality (Kimura et al., 2011). For both modalities, simplistic sensory violations, like infrequent tones in audition or color changes in vision produce such (MMN) effects. That such an effect may arise from a sensory violation on an abstract level was only known for the auditory modality (Frangos, Ritter, & Friedman, 2005). Therefore, our finding of a potential vMMN component elicited by a violation of an abstract prediction in vision (violation of semantic prediction on visual stimuli) is of special interest. Our finding contributes not only to the evidence of sensory prediction, but also to the general MMN discussion (Katrina Quinn, p. 16–17).

Interestingly, with each successive character in a sequence, we found a trend of an increasing amplitude in its respective VEP (1st component), even though this trend was not significant. Possibly, it reflects an increase in the facilitation of early visual processing (e.g. through prediction), because stimuli occur in sequence. But this effect was not strong enough to be reliably detected in all of our subjects. The slight trend might reflect (a) a general effect of presenting stimuli in sequence, or (b) a creation of a prediction, or (c) reflect a confound. This confound could arise due to the color change of the rectangle during response epoch and/or due to the variable length of the test interval.

Further studies could clarify this possible facilitation effect, by repeating our experiment. The amplitude effects we just described could be compared with the amplitude of the early VEPs to a stream of letter stimuli in non-alphabetical order. If the amplitude effects were found for the latter condition, this would suggest that

these effects come from a general facilitation of successively presented sensory stimuli. If no effect on amplitudes were found, “it may provide a starting point for further insights into how predictions influence sensory processing in visual cortex” (Katrina Quinn, lab report, p. 17).
“One problem with our design is that we were unable to analyze certain aspects of both the first and second components of the ERP. This was due to the change in color of the rectangle during the response phase of the experiment, which produced an early VEP as well. This confounded VEPs related to the last predicted/unpredicted letter, and meant that we could not make meaningful comparisons of amplitude for the first vs. the last letter across all components of the VEP. In order to improve upon this, adaptations of the paradigm should aim to devise a way in which the participant could be alerted to the presence of the response phase, without the addition of confounding VEPs (but see the problems that emerge from such attempts, as were present in Designs #03–05). […] [A] suggestion would be to design a new experiment in which a temporal judgement of the onset of a stimulus is made for each and every stimulus along the stimulus train, all without being confounded by the next stimulus. Whilst this may involve a complex design, the number of comparisons that could be made would be greatly beneficial to clarifying the effects discussed in this report” (K. Quinn, pp. 17–18). However, the confounding effect should not influence the comparison of predictive versus non-predictive condition. This is because the confound is consistent across the two conditions.

To warrant sufficient data quality for MEG analysis, it would have been desirable if we had been able to reproduce the behavioral results from the IR-laboratory environment of Design #06 in the MEG environment as well. However, the behavioral effect was only marginally significant for the dataset acquired in MEG environment. A robust reproduction of the behavioral effect in the MEG environment and a subsequent analysis of related MEG data might already be a major improvement. If behavioral data would map directly into MEG data the way we expected, it is no wonder that there is no effect in the present MEG data when the related behavioral effect is statistically absent. Although we found no apparent difference in the timing of early sensory processing for unpredicted vs. predicted stimuli, we found an MMN-like signal. We interpreted this MMN-like signal as a correlation for prediction. According to the MMN-like signal, postdiction is likely to play a role. Still, we cannot provide compelling evidence. Because of the aforementioned missing difference in timing of sensory processing, we are not able to make a final decision regarding whether the underlying mechanism (for MMN-like signal) is predictive or postdictive or an interplay of both.

Our results extend the view on perception and its contribution in delay compensation. Behavioral data support the idea that delays may be compensated at the perceptual level, at least in visual modality. However, the underlying mechanism that mediates this effect remains somewhat unclear. According to the electrophysiological data from our MEG study, *post*diction is likely to play an additional role (as shown by the presence of a MMN-like signal). A purely predictive mechanism seems unlikely. A combination of prediction and postdiction or another, yet unknown mechanism might underlie the physiological and behavioral findings. Interestingly, according to Kimura et al. (2011), the most plausible approach to vMMN is the *pre*diction-error account (not an MMN), as illustrated in Figure 36 of the present thesis.

Delimitation from other temporal distortions

We want to delimit the present study from the following temporal distortions. Wherever possible, we also want to discuss how some of these distortions might have interfered with the non-successful designs of this study. However, these interference-approaches are speculative.

Representational Momentum

For a moving target, its memorized final position is often displaced in the direction of target motion. This phenomenon is known as representational momentum (RM) (Hubbard, 2005). Like the FLE, RM is a displacement error. In RM, a moving object is a prerequisite. The fact that motion is a prerequisite for RM delimits the RM research from the aspect of time-perception that we focus on. In the present study, motion is not displayed, nor do subjects move their head and eyes. Hence, although RM is related to time perception, it is not related to the kind of time perception and prediction addressed in the present study.

Multiple approaches to RM exist, ranging from low-level perceptual mechanism, to high-level cognitive mechanism; they need to integrate a wide range of empirical data (Nijhawan & Khurana, 2010).

Prior entry

Prior entry is a bias in temporal judgement tasks. "The object of attention comes to consciousness more quickly than the objects which we are not attending to," as was first described by Titchener as one of the fundamental laws of attention (cited according to Spence & Parise, 2010). According to this view, paying attention to a stimulus leads to an earlier arrival of sensory signals arising from this stimulus at some critical brain center, compared to sensory signals arising from an unattended stimulus. Hence, if attention is directed toward a stimulus, it is proposed to lead to an "acceleration" of information processing in cortical sensory pathways. McDonald and colleagues argue against the attention-approach to the prior entry phenomenon, because they found no evidence for the widespread view that attended objects are transmitted more rapidly through the visual system than unattended objects (McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2005). No matter what the underlying mechanism for prior entry is, there is experimental evidence on behavioral level that points of subjective simultaneity are significantly different between conditions with attended target stimuli and unattended target stimuli (McDonald et al., 2005).

We did not intend to investigate prior entry, but prior entry might interfere with our results. However, attention was equally distributed across conditions. The display consists of a stream of objects, each of which is presented separately and at identical positions. Prior entry would require a comparison of two conditions presented simultaneously, e.g. two streams of objects displayed at the same moment in two different locations. Attention could then be directed to one stream—or one single object within this stream—and therefore attention could be drawn away from the other stream. In the present study's design(s), attention cannot be drawn away in the same way as discussed in the context of prior entry.

The aforementioned focuses on prior entry in context of spatial attention. Let's assume that attention is drawn towards new, surprising stimuli. In this case one could

argue that if attentionis shifting towards a new, unpredictable stimulus, it could lead to prior entry as well. On the one hand, this is speculative. On the other hand, even if there was such a shift (in our case attention would shift towards the non-predictive character), this effect would diminish the predictive effect. In other words, it would work against the predictive effect.
Taken together, prior entry in the context of spatial attention does not interfere with the present study's design. In the context of (speculative) attention shift toward surprising stimuli, it would reduce the predictive effect. Therefore, prior entry is not suitable to cause false positive results in our study.

Prior entry might have caused or contributed to a side effect in Design #04 (auditory cueing). There was a surprising change in PSE values (~800 ms) in Design #04 (auditory cueing). The auditory cueing might have drawn the subjects' attention away from the visual stimulus. The character subsequent to the auditory-cued character might be perceived late, since the character did not receive the subject's full attention. Therefore, the last regular ISI might appear longer than the regular ISIs before. Remember, subjects were instructed to judge the duration of ISIs and not the duration of displayed characters. Therefore, the variable ISI, which was the test ISI, needed to physically expand in time, to be perceived as equally long with standard ISIs. When attention shifts back to vision to the full extent (the auditory cue was only used once per trail for the penultimate character), the last character might have been perceived relatively earlier. Hence, the ISI had to be physically shorter in order to be perceived as equally long as that preceding the penultimate letter. This would result in lower PSE values than the actual ISI, as observed in this experiment. Whether the lower PSEs were induced by prior entry in the aforementioned way is clearly highly speculative.

Hazard rate of time

The 'hazard rate of time' is an internal representation of elapsed time and the probability that a certain signal/behaviorally relevant event is about to occur—given that it has not occurred yet. The concept of a hazard rate of time was investigated in rhesus monkeys (Janssen & Shadlen, 2005). Temporal expectations are known to modulate perception (Nobre, Correa, & Coull, 2007).The neural mechanism underlying the encoding of elapsed time is the object of research. Humans, as well as animals, need a sense of elapsed time to e.g. deduce casual regularities or to predict salient events. This is why this concept is of interest for our study.

To our knowledge, the underlying mechanism for a hazard rate of time remains unclear. Whatever the underlying (neural) mechanism might be, the hazard rate of time should not interfere with our experiment design. This is because of equal temporal expectations across conditions. Temporal expectations might modulate the temporal percept—e.g. the fourth character may differ from the seventh character. Since sequence length is equally long across conditions, temporal expectation/hazard rate should not interfere with our experiment design.

Saccadic Chronostasis

Chronostasis is a temporal distortion in which the “apparent duration of stimuli following saccades“ is increased (Morrone, Ross, & Burr, 2005). This phenomenon is thought to “compensate perceptually for the time lost during saccadic suppression” (Morrone et al., 2005).

Saccadic eye movements are known to cause a subjective temporal lengthening of a visual stimulus. A widespread notion/interpretation of this effect is that the onset of the postsaccadic stimulus is retrospectively antedated to a moment just before the saccadic movement was initiated. When subjects make a rapid eye movement toward a new visual target, they overestimate the duration of that new target. Accordingly, the magnitude of saccadic chronostasis is greater for longer saccades, whereas the duration of the visual stimulus has no influence (Yarrow, Haggard, Heal, Brown, & Rothwell, 2001).

We lack visual experience during both saccades and blinks. During saccadic eye movements, we do not reliably perceive a shift in space of stationary objects. They also do not smear out. During blinks, we cannot see the world, because the eyes are closed. Nevertheless, we do not experience ‘nothing’ or a black frame during blinks and saccades. Chronostasis might partially explain perceptual continuity despite saccades and perhaps also despite blinks.

Since our study requires temporal judgments about visual stimuli, we wanted to avoid any interference with known temporal distortions like chronostasis. To prevent chronostasis from masking the ‘real’ predictive effect, we monitored eye movements.

Data collected during and around saccadic eye movements were excluded from analysis. We traded blinks in the same way as saccades. We expected the highest impact of temporal distortions caused by saccades and blinks (chronostasis) during the last phase of each stimulus epoch, when the last standard interval and test interval are presented. This is when the subjects had to make their temporal judgment. We did not check for eye movements at the beginning of the stimulus epoch. If there were eye movements, they should affect the time judgment less, because there are several standard intervals presented based on which the subject makes her/his temporal judgment.

Although chronostasis is well-known, in some studies, eye movements and blinks were not monitored, such as in the study by Eagleman and Sejnowski on postdiction and visual awareness. However, we did monitor eye movements and blinks in our study and excluded “contaminated” data from our analysis.

Temporal distortions of perceived stimulus duration.

Events sometimes appear to last longer or shorter than events of equal length. Experimental evidence exists for both domains, audition and vision. I will here refer to visual effects only. Three temporal distortions are of particular interest—the “magnitude,” “oddball,” and “debut” effects. Their commonality is an effect on perceived *stimulus duration* based on predictability. This commonality also delimits the following effects from chronostasis.

Magnitude effect.

Stimulus magnitude has an effect on perceived duration this effect is termed magnitude effect. In an experiment with numbers, e.g. a “7” is perceived to last longer than a “1,” although they are presented for an equally long duration (Pariyadath & Eagleman, 2007). In addition, magnitudes in temporal and non-temporal dimensions seem to be interdependent, implying that generalized and abstract components exist in magnitude representation (Xuan, Zhang, Chen, He, & Chen, 2007).

I argue that a magnitude effect cannot account for my results. We aimed to avoid an interference with the magnitude effect. Therefore, we decided against the use of numbers as stimuli. Still, we aimed to use stimuli which were inherently predictive. This is also why we included only subjects raised with the latin alphabet as their first language. It is not clear whether progression in alphabetical order can elicit a magnitude effect. Even if this was the case, this effect should equal out across both conditions and trials. The sequences had start-characters chosen at random and were of different length (four to eight characters). The last character in the NP-condition was chosen from both letters that were preceding and those that were following it. Therefore NP characters were not systematically chosen from, for instance, the subsequent part of the alphabet, which might have possibly evoked a result due to a magnitude-effect.

Taken together, the magnitude effect should not interfere with the present study design. Although it is not known whether a subsequent sequence of characters is suitable to evoke a magnitude effect, such an effect would balance out across both conditions.

Oddball effect.

Studies that investigated the effects of predictability on perceived *duration* found that an oddball appearing in a sequence of expected events seems to last longer. In other words, the stimulus expands subjectively in time. The odd stimulus is overestimated by up to 50%, although oddball and regular stimulus are physically presented for an equally long duration. This is the same for both auditory and visual modality (Pariyadath & Eagleman, 2007; Tse et al., 2004). In vision, the oddball effect was often investigated by visual stimuli, with the same edges and shapes for repeated stimuli. However, to my knowledge, no one has looked at the effect of event predictability on perceived onset.

The oddball effect is not suitable to systematically interfere with the key results of the present study. First, as mentioned before, subjects were instructed to judge the duration of ISIs and therefore the perceived *onset*, not the duration of displayed characters mattered. The oddball effect refers to *stimulus duration*, not to duration of inter stimulus intervals. (To my knowledge there is no evidence in literature for interplay of perceived stimulus duration and stimulus onset in context of oddball effect.) Second, the oddball effect is often explained via an attention approach, although the attentional approach is not suitable to fully explain the effect. Emotionally salient oddballs are supposed to attract more attention. Since emotionally salient oddballs do not show an increased oddball effect, attention cannot fully explain the temporal distortion of an oddball (Pariyadath & Eagleman, 2007). In any case, in the present study, attention is equally distributed across conditions, as discussed before (in the context of prior entry).

Debut effect.

Like an oddball, the first stimulus in a sequence of repeatedly presented stimuli appears to last longer than the following presentations. This is known as the debut effect (Pariyadath & Eagleman, 2007) and is pretty similar to the oddball effect. An oddball within a sequence is as unpredictable as a stimulus at the beginning of a sequence. One could question for both effects, oddball and debut, if they interfere with the current study design. As mentioned before, both illusions have an effect on perceived stimulus duration, whereas in our study the onset of a stimulus was crucial. To my knowledge, there is no evidence in the literature for the interplay of perceived stimulus duration and stimulus onset in the context of oddball effect and/or debut effect.

Still, with respect to the NP condition, the last, non-predictable character was perceived earlier than the last character of the P condition, this would lead to a systematic difference between both conditions. But such a systematic difference would diminish a predictive effect. Neither the debut effect nor the oddball effect can magnitude,oddball, is known to have an effect on perceived duration of visual stimuli and is related to the predictability of the stimuli. The present study focuses on the stimulus onset.

Even if the perceived duration of presented characters in our study was affected, as a side effect the ISI could be affected as well. We assume that, if one temporal perception was compressed, the other temporal perception needed to expand. Otherwise, temporal perception would uncouple from physical time after some repetition cycles of stimulus–ISI–stimulus–ISI and so on. Such an uncoupling seems unlikely and does not correspond to our daily experiences in interactions with our environment.

In case the NP character was perceived earlier than the P character (last character of P condition), this led to a systematic difference between both conditions. But such a systematic difference would diminish a predictive effect, not increase it. None of the aforementioned effects is suitable to cause false positive results in our study (similar to prior entry). As argued above, this could possibly reduce the predictive effect in our study, but not lead to false positive results.

Even if the perception of ISIs was distorted by the aforementioned, it led to no systematic difference between predicted and unpredicted conditions. Duration of displayed objects (characters) and ISIs was identical across conditions. The only difference between conditions was the last character, which might be predictive (P) or non-predictive (NP). Therefore, any difference found in PSE values should reflect the predictability of the last character.

Eagleman and Pariyadath point out that both the debut and the magnitude effect have a direct parallel in electrophysiological studies of repetition. A reduced neural response to repeated stimuli is known as repetition suppression. In response to repeated stimuli presentation, the firing rate of neurons in higher cortical areas diminishes. In addition, amplitudes of ERP signals (in rhesus monkeys) decrease in response to the second presentation of a stimulus (Pariyadath & Eagleman, 2007). It is suggested that the suppression of neural responses to repeated stimuli allows the system to save resources and thus reflect a more efficient presentation. Alternatively, it could allow novel (and perhaps ecologically more relevant) stimuli to capture processing resources more easily. No matter what the underlying mechanism of repetition suppression might be, it seems to directly map onto behavioral data—conditions that lead to a reduced neural response on the physiological level correlate with those causing a shortened perceived duration (Pariyadath & Eagleman, 2007). If this mapping was meaningful and correct, the neural response in an abstract sequencelike the alphabetical order in our study—should be suppressed and lead to a reduction in perceived duration.

Conclusion and Outlook

Our findings demonstrate that predictable visual stimuli are in fact perceived earlier than non-predictable stimuli at the behavioral level. This suggests that even the perceptual system compensates for delays in sensory information processing, allowing us to establish a timely percept of our environment.

At the electrophysiological level, we aimed to shed light on the underlying neuronal mechanisms of an earlier perception of predicted vs. non-predicted stimuli. The analysis of MEG data suggests that participants generated predictions. This was indicated by a signal similar to vMMN, namely a signal difference for the non-predicted as compared to the predicted stimulus. However, we found no evidence for predicted stimuli being processed earlier than unpredicted stimuli. Hence, the psychophysical effect underlying our subjective temporal benefit in perceiving predictive stimuli is likely based on a postdictive process, which compensates for neural delays retrospectively. Ultimately, it remains open whether the earlier perception of predictive vs. non-predictive stimuli is mediated through sensory prediction, through an interplay of prediction and postdictive perceptual evaluations, or through prediction and some yet unknown delay compensation mechanism. Further research is required to examine the role of postdiction vs. prediction and the possibility that visual components could be affected by predictions.

References

Papers and Books

Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, *26*(5), 340–354. https://doi.org/10.3758/BF03204158

Blakemore, S.-J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Experimental Brain Research. Experimentelle Hirnforschung.*, *153*(2), 239–245. https://doi.org/10.1007/s00221-003-1597-z

Cavanagh, P. (1997). Predicting the present. *Nature*, *386*, 19–21. https://doi.org/10.1038/35002708

Changizi, M. a. (2001). “Perceiving the present” as a framework for ecological explanations of the misperception of projected angle and angular size. *Perception*, *30*(2), 195–208. https://doi.org/10.1068/p3158

De Valois, R. L., & De Valois, K. K. (1991). Vernier acuity with stationary moving gabors. *Vision Research*, *31*(9), 1619–1626. https://doi.org/10.1016/0042-6989(91)90138-U

Dreher, B. Y. B., Fukada, Y., & Rodieck, R. W. (2006). ANATOMICAL SEGREGATION OF CELLS WITH X-LIKE termed Type Illy ., (1976), 433–452.

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science (New York, N.Y.)*, *287*(5460), 2036–2038.

Frangos, J., Ritter, W., & Friedman, D. (2005). Brain potentials to sexually suggestive whistles show meaning modulates the mismatch negativity. *Neuroreport*, *16*(12), 1313–1317. https://doi.org/10.1097/01.wnr.0000175619.23807.b7

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.97.1.35

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin {&} Review*, *12*(5), 822–851. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16524000

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, *9*(4), 304–313. https://doi.org/10.1038/nrn2332

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience*, *8*(2), 234–41. https://doi.org/10.1038/nn1386

Jeannerod, M., Kennedy, H., & Magnin, M. (1979). Corollary discharge: Its possible implications in visual and oculomotor interactions. *Neuropsychologia*, *17*(2), 241–258. https://doi.org/10.1016/0028-3932(79)90014-9

Kandel, E. R., Schwartz, J. H., Jessel, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2011). *Principles of neural science* (5th ed.). New York: McGraw-Hill Medical Publishing Division.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr Opin Neurobiol*, *9*(6), 718–727. https://doi.org/10.1016/S0959-4388(99)00028-8

Kimura, M., Schröger, E., Czigler, I., Kimura, M., Schröger, E., & Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *Neuroreport*, *22*(14), 669–673.

McDonald, J. J., Teder-Sälejärvi, W. a, Di Russo, F., & Hillyard, S. a. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience*, *8*(9), 1197–202. https://doi.org/10.1038/nn1512

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience*, *8*(7), 950–4. https://doi.org/10.1038/nn1488

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*(1), 6–18.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2007.04.026

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, *370*, 256–257.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive Sciences*, *6*(9), 387. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12200181

Nijhawan, R. (2008). Visual prediction: psychophysics and neurophysiology of compensation for time delays. *The Behavioral and Brain Sciences*, *31*(2), 179-

98-239. https://doi.org/10.1017/S0140525X08003804

Nijhawan, R., & Khurana, B. (2010). *Space and time in perception and action*. Cambridge: Cambridge University Press.

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–70. https://doi.org/10.1016/j.conb.2007.07.006

Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PloS One*, *2*(11), e1264. https://doi.org/10.1371/journal.pone.0001264

Ratliff, F., & Hartline, H. K. (1959). The responses of Limulus optic nerve fibers to patterns of illumination on the receptor mosaic. *The Journal of General Physiology*, *42*(6), 1241–1255. https://doi.org/10.1085/jgp.42.6.1241

Schmolesky, M. T., Wang, Y. C., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *JOURNAL OF NEUROPHYSIOLOGY*, *79*, 3272–3278.

Schwartz, G., Harris, R., Shrom, D., & Berry, M. J. (2007). Detection and prediction of periodic patterns by the retina. *Nature Neuroscience*, *10*(5), 552–554. https://doi.org/10.1038/nn1887

Spence, C., & Parise, C. (2010). Prior-entry: a review. *Consciousness and Cognition*, *19*(1), 364–79. https://doi.org/10.1016/j.concog.2009.12.001

Stefanics, G., Kremlácek, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in Human Neuroscience*, *8*, 666. https://doi.org/10.3389/fnhum.2014.00666

Stöhr, M., & Bach, M. (2005). *Evozierte Potenziale: SEP, VEP, AEP, EKP, MEP* (4th ed.). Heidelberg: Springer Medizin.

Taylor, M., & Creelman, C. (1967). PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America*, *41*, 782–787. https://doi.org/10.1121/1.1910407

Toga, A. W., & Mazziotta, J. C. (2002). *Brain mapping: The methods* (2nd ed.). Amsterdam and and Boston: Academic Press. Retrieved from http://www.worldcat.org/oclc/52594824

Tresilian, J. R. (1993). Four questions of time to contact: a critical examination of research on interceptive timing. *Perception*. https://doi.org/10.1068/p220653

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics*, *66*(7), 1171–89. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15751474

Wirxel, B., & Lindner, A. (2012). *"Predictable visual stimuli are perceived earlier than unpredictable events "* (Society for Neuroscience 2012; Conference Poster; New Orleans, USA)

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology : CB*, *11*(18), R729-32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12546789

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3 Suppl*(november), 1212–1217. https://doi.org/10.1038/81497

Xuan, B., Zhang, D., Chen, X., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, *7*, 1–5. https://doi.org/10.1167/7.10.2.Introduction

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, *414*(6861), 302–5. https://doi.org/10.1038/35104551

Online resources

http://www.vislab.ucl.ac.uk/cogent.php
(free psychophysics software, 'Cogent', Version: Cogent2000v1.30)

http://www.r-project.org/
(free statistics software, 'R', Version: R i386 2.15.0, date: 29. August 2012, 16:01:02)

http://pixabay.com/de/audio-ton-lautsprecher-button-150191/
(Sound Symbol used in Design #04. 21. May 2014. 15:11. License: CC0 1.0 Universell (CC0 1.0) Public Domain Dedication)

http://www.ru.nl.neuroimaging/fieldtrip (software for MEG data analysis)

https://www.mendeley.com/ (free reference managing software)

Appendix

Tabular overview of experiment designs

Design	Characteristics	Substantial novelty
#01	- n=12 (pilots) - Hz: 100 - sequence length: 4-8 - character display: 300ms (each) - standard interval: 1000ms (each) - start values: 1500ms; 1200ms; 300ms (for each condition) - initial step size: -320; -320; 320 (for each condition) - eye tracking: yes - 'dual task': Yes, used for post-hoc correction after the experiment - Strategy: PEST P: 25%, 50%, 75% (3x30 trials) NP: 25%, 50%, 75% (3x30 trials) Wald-constant =1 - share: Equal share of P&NP; - trials: 180 (2x 90) - 'test interval': varied, according to Pest Strategy. stimulus name: stim_TimePredictionFinalVersion1.m (folder: Time Prediction - Teil I)	None
#02	- n= 34 - Hz: 100 - sequence length: 4-8 - character display: 300ms (each) - standard interval: 1.000 ms - eye tracking: yes, offline - 'dual task': yes, online control Pretest ('old' – used until Sub no 08) - strategy: PEST - Wald-constant: 1.0 - share: equal share of P&NP - trials: 50	- Two experimental parts: Pre-test & Main experiment - Feedback on control task (alphabetical order)

	- 'test interval': varied, according to PEST strategy - start values: 2000; 400 - initial step size: 1280; -1280 pre-test ('new pretest – used from sub no09') - strategy: PEST - Wald-constant: 0.75 - share: equal share of P&NP - trials: 90 - 'test interval': varied, according to PEST strategy - start values: 1200ms; 1700ms; 700ms - initial step size: 1280; 1280; -1280 main experiment - strategy: MCS 11 values: pPSE ± 250 ms in 50ms steps 11 values*10 repetitions (110 trials) for each condition - share: equal share of P & NP - trials: 220 (110 P & 110 NP) - 'test interval': fixed for each subject individually (11 values) stimulus names: stim_timePredictionPRETEST_rough.m (old pretest) stim_TimePredictionPRESTST_new.m (new pretest) stim_TimePredictionMCS.m (folder: WorkingVersion – Time Pred Teil II)	
#03	- n= 6 - Hz: 100 - sequence length: 4-8 - character display: 300ms (each) - standard interval: 1.000 ms - eye tracking: yes, online - 'dual task': yes, online control pre-test - strategy: PEST - Wald-constant: 0.75- share: equal share of P&NP	- online Eyetracking including feedback on eye movements - fixation cross - last character is oviously the last character for both conditions - circles during response phase

	- trials: 90 - 'test interval': varied, according to PEST strategy - start values: 1200; 1700; 700 - initial step size: 1280; 1280; -1280 main experiment - strategy: MCS 11 values: pPSE ± 250 ms in 50ms steps 11 values*10 repetitions (110 trials) for each condition - share: equal share of P & NP - trials: 220 (110 P & 110 NP) - 'test interval': fixed for each subject individually (11 values) stimulus names: stim_TimePredictionMCS_III_Red.m und stim_TimePredictionPRE-TEST_III-Red.m (folder: Working Version –Time Prediction Teil III)	
#04	- n= 12 - Hz: 100 - sequence length: 4-8 - character display: 300ms (each) - standard interval: 1.000 ms - eye tracking: yes, online - 'dual task': yes, online control pre-test - strategy: PEST - Wald-constant: 0.75 - share: equal share of P&NP - trials: 90 - 'test interval': varied, according to PEST strategy - start values: 1200; 1700; 700 - initial step size: 1280; 1280; -1280 main experiment -strategy: MCS 11 values: pPSE ± 250 ms in 50ms steps 11 values*10 repetitions (110 trials) for each	- sound as label - no frames - pause - response epoch with words

	condition - share: equal share of P & NP - trials: 220 (110 P & 110 NP) - 'test interval': fixed for each subject individually (11 values) stimulus names: stim_TimeMCS_4.m stim_TimePRETEST_4.m (folder: Time Prediction - Teil 4)	
#05	- n= 7 - Hz: 100 - sequence length: 6 - character display: 300ms (each) - standard interval: 1.000 ms - eye tracking: yes, online - 'dual task': yes, online control pre-test - strategy: PEST - Wald-constant: 0.75 - share: equal share of P&NP - trials: 90 - 'test interval': varied, according to PEST strategy - start values: 1200; 1700; 700 - initial step size: 1280; 1280; -1280 main experiment - strategy: MCS 11 values: pPSE ± 250 ms in 50ms steps 11 values*10 repetitions (110 trials) for each condition - share: equal share of P & NP - trials: 220 (110 P & 110 NP) - 'test interval': fixed for each subject individually (11 values) stimulus names: stim_TimeMCS_5.m stim_TimePRETEST_5.m stim_TimeTRAINING_5.m	-no Click-sound -sequence length fixed to 6 characters. - pause remains unchanged

#06 @ IR	- n= 8 - Hz: 100 - sequence length: 4-8 - character display: 10 ms (one frame each ch.) - standard interval: 1000 ms - eye tracking: yes, online - 'dual task': yes, online control pre-test - strategy: PEST - Wald-constant: 0.75 - share: equal share of P&NP - trials: 90 - 'test interval': varied, according to PEST strategy - start values: 1700; 700 - initial step size: 1280; -1280 main experiment - strategy: MCS 11 values: pPSE ± 250 ms in 50ms steps 11 values*10 repetitions (110 trials) for each condition - share: equal share of P & NP - trials: 220 (110 P & 110 NP) - 'test interval': fixed for each subject individually (11 values) stimulus names: stim_TimePredictionTRAINING_6NoE stim_TimePredictionTRAINING_6 stim_TimePredictionPRESTST_6 stim_TimePredictionMCS_6	-Character display duration: one frame (10 ms) -------------------- back to #02: - response with rectangles - <u>no</u> fixation cross -------------------- Changes compared to #02: -online eye tracking
#06 @ MEG	- n=11 - Hz: 60 - sequence length: 4-8 - character display: 16,67 ms (one frame) (each) - standard interval: 983,3ms (1000ms-16,67ms) - eye tracking: yes, online - 'dual task': yes, online control	Changes due to technical constrains: -changed Hz rate (60 Hz) - Character display duration: one frame (16,67 ms) -changed duration of standard intervals (983,3ms)

	pre-test - strategy: PEST - Wald-constant: 0.75 - share: equal share of P&NP - trials: 90 - ‘test interval’: varied, according to PEST strategy - start values: 1700; 700 - initial step size: 1280; -1280 main experiment - strategy: MCS 11 values: pPSE ± 250 ms in 50ms steps 11 values*10 repetitions (110 trials) for each condition - share: equal share of P & NP - trials: 220 (110 P & 110 NP) - ‘test interval’: fixed for each subject individually (11 values) stimulus names: stim_T6_TRAINING_NoE_MEG stim_T6_TRAINING_6_MEG stim_T6_PRETEST_6 stim_T6_MCS_6_MEG	- marker for MEG-data acquisition -session split in 12 min blocks Changes due to analysis: -marker for analysis of MEG data (not visible to the subject) -baseline period (983,3ms) -------------------- back to #02: - response with rectangles - no fixation cross -------------------- Changes compared to #02: -online eye tracking

List of figures

List of tables

Visual angle

Calculation of visual angle as follows: Viewing distance from eye to monitor surface was approximately 57 cm. Screen resolution (IR lab) was 1280*1024 pixel (width*height). Screen dimension were 40 cm*30 cm (width*height). Calculation was with horizontal direction. X- and Y-dimensions of pixel size were usually the same. One centimeter on screen equaled one-degree visual angle at 57cm viewing distance. For an object centered in the middle of the screen, as it was the case for all my experiments, calculation is as follows:

$\tan \alpha$ = opposite leg / adjactant leg
$\tan \alpha$ = 1 [cm] / 57 [cm] = 0.0175
α = 1.005°

40 [cm]/ 1280 [pixel] = 0.0313 [cm/pixel]
1 [cm] = 1°
0.0313 [cm/pixel] => 0.0313° (This value is required for eye analysis (variable name: pix2deg).)

Conference contribution

Conference poster: Predictable visual stimuli are perceived earlier than unpredictable events (910.04/CCC5); B. WIRXEL, A. LINDNER; Society for Neuroscience, New Orleans, US, Oct. 2012

Conference poster: Predictable visual stimuli are perceived earlier than unpredictable events; B. WIRXEL, A. LINDNER; Timely; Corfu, Greece, Feb 2013

Conference poster: It was (not) me: Causal Inference of Agency in goal-directed actions. Tobias F. Beck' Carlo Wilke, Barbara Wirxel, Dominik Endres, Axel Lindner& Martin A. Giese (http://precedings.nature.com/documents/5858/version/1)

Acknowledgement

This work was supported by the Bernstein Center for Computational Neuroscience (BMBF; FKZ: 01GQ1002) and CIN, Tübingen. A travel stipend to the Society for Neuroscience—Meeting 2012 was provided by the Graduate Training Centre of Neuroscience, Tübingen, Germany. A travel stipend for the TIMELY conference 2013 (Time In MEntaL activitY) was provided by COST.

Declarations/ Erklärung

Erklärungen nach § 7, Abs. 2, der Promotionsordnung

Ich erkläre hiermit,

- dass ich mich bisher keinen Promotions- oder entsprechenden Prüfungsverfahren unterzogen und diese abgebrochen oder abgeschlossen habe.

- dass die vorgelegte Dissertation noch nie ganz oder teilweise als Dissertation oder sonstige Prüfungsarbeit eingereicht worden ist und auch nicht ganz oder teilweise veröffentlicht worden ist.

  Im Fall von § 9 Abs. 2 ist eine vom Promotionskomitee bestätigte Erklärung des Doktoranden hinsichtlich aller Veröffentlichungen, bei denen mehrere Autoren mitgewirkt haben, beizufügen (Darstellung des Eigenanteils).

- dass ich die zur Promotion eingereichte Arbeit mit dem Titel

  "Human Time Perception –
  Predictable visual stimuli are perceived earlier than unpredictable events"

  selbstständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

- dass ich bisher weder strafrechtlich verurteilt, noch Disziplinarmaßnahmen und anhängigen Straf- und Disziplinarverfahren unterzogen worden bin , soweit die Auskunftspflicht nicht durch § 51 des Bundeszentralregistergesetzes ausgeschlossen ist,

- dass mir die Gelegenheit zum vorliegenden Promotionsverfahren nicht kommerziell vermittelt wurde und dass ich keine Organisation eingeschalten habe, die gegen Entgelt Betreuer für die Anfertigung von Dissertationen sucht oder die für mich die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen ganz oder teilweise erledigt. Ich erkläre des weiteren, dass mir die Rechtsfolge der Inanspruchnahme eines gewerblichen Promotionsvermittlers und die Rechtsfolge bei Unwahrhaftigkeiten in dieser Erklärung (Ausschluss der Annahme als Doktorand, Ausschluss der Zulassung zum Promotionsverfahren, Abbruch des Promotionsverfahrens und Rücknahme des erlangten Grades gemäß § 25) bekannt sind.

Tübingen, den 07.08.2017 ……………………………………….

Datum Unterschrift

Danksagung

Schreiben musste ich diese Arbeit zwar allein – doch an ihrem guten Gelingen waren Menschen beteiligt, denen ich meinen Dank aussprechen möchte.

An erster Stelle, die anderen mögen es mir nachsehen, danke ich meinen Doktorvater, PD Dr. Axel Lindner. Für die Betreuung des Themas, den Enthusiasmus, mit dem er die Forschung betrachtet und die er meiner eigenen Skepsis entegegensetzte; für Unterstützung und Hilfe bei Fragen zum experimentellen Design, Eyetracking und programmiertechnischen Herausforderungen, um nur einige zu nennen. Das habe ich nie als selbstverständlich angesehen. Den Zusammenhalt innerhalb des Labors – mögen die Forschungsthemen auch verschieden, und eine räumliche Trennung gegeben sein - das freundliche Interesse untereinander, daran hat Axel großen Anteil. Herzlichen Dank.

Prof. Christoph Braun möchte ich für den Zugang zum MEG-Center, die freundliche Unterstützung bei der MEG-Studie und das Interesse an meiner Arbeit danken. Ebenfalls bedanken möchte ich mich für sein Engagement im Advisory Board.

Prof. Uwe Ilg danke ich für die wohlwollende Begleitung des Forschungsvorhabens, diverse Anregungen und Hinweise und sein Einsatz im Advisory Board. Besonderer Dank gilt dem Hinweis auf den *einen* frame. Jeder kann nur mit seinem eigenen Kopf denken – sein aufmerksamer Hinweis brachte mich auf die richtige Spur.

Darüber hinaus gilt mein Dank dem gesamten Nodlab für Diskussionen, Anregungen und den Zusammenhalt innerhalb des Labors. Jeder war hat mal für den anderen ein Setup getestet oder als Proband teilgenommen, weil es für jeden/n selbstverständlich war, sich die Zeit zu nehmen. Das ist eine schöne Erinnerung. Besonders bedanken möchte ich mich bei Melanie Höller, die mich an ihrem Statistik-Wissen teilhaben ließ. Danke für die fachlichen Diskussionen und motivierende Worte, wenn es mal schwierig war. Christoph Budziszewski danke ich besonders für Tipps und Tricks beim Programmieren in MatLab, Manuel Roth für fachliche Unterstützung und Rat.

Ein besonderer Dank geht an Katrina Quinn, die als Masterstudentin im Rahmen einer 6-wöchigen lab rotation die Analyse der MEG-Daten durchführte.

Ein dank natürlich auch, an die beiden unbekannten *Models*, die sich für die vorliegende Arbeit in den beiden Setups ablichten ließen.

Außerhalb des wissenschaftlichen Umfeldes gab es Menschen, denen ich danken möchte. Meiner guten Freundin, Stefanie Wilkes, die stets ein offenes Ohr für mich hatte, mich moralisch unterstütze und an das Gelingen der Arbeit glaubte. Dr. Ulrike Banning, meiner Biologen- Freundin, die die Höhen und Tiefen einer Promotion kennt, und in der ich Verständnis fand. Meiner Schwester, dass sie mich oft moralisch unterstütz und an das Gelingen der Arbeit geglaubt hat.

Einen ganz besonderen Platz nimmt Katrin Kutscheidt für mich ein, die leider viel zu früh von uns gegangen ist. Danke für die gute gemeinsame Zeit, die Gespräche, die Besuche in Mannheim. Denke ich an die Zeit im Hertie, denke ich an Dich.