Michael H. Herzog

The Temporal Microstructure of Conscious Perception

1. Introduction

Intuitively, there is no problem with time. We feel that time passes by in a smooth, continuous fashion. A diver jumps from a cliff, and we see her trajectory at each single moment in time. Continuous motion in the external world elicits a continuous stream of conscious percepts. However, when dissecting the problem, happy intuition quickly ends up in puzzling questions, which are far from being answered even though the discussion has started more than 1500 years ago.

St. Augustine was driven to despair when thinking about time. Since the »now« is an infinitesimally[1] short moment of time, we can be conscious only at these short moments. However, how can we then perceive motion and change, which, by definition, need to rely on the past and, potentially, the future? St. Augustine proposed that each conscious percept is restricted to a moment and change, and motion percepts are derived from memory:[2] We are conscious of the current state of the world and may remember past states, which can give rise to the »conclusion« that things have changed. A classic example is that you may leave your home and see the neighbors' trash bin at the walkway. When you return to your home later, the trash bin may be at a different position, and you conclude that it has been moved. Whereas such inferences indeed play a role in cognition, they do not

1 I am freeing myself from hermeneutic philosophy, sticking to an interpretation, based on my very restricted knowledge of classic texts, formulated in a modern science language; the idea is to mainly lay out the main problems, as I see them, within a historically disguised context; for example St. Augustine did not use the word »infinitesimal«, which here is used in its mathematical meaning, e.g., when time (t) is represented on the x-axis with real numbers.

2 Following Dainton (2018), models of this type are nowadays called cinematic models. See Figure 1.

seem to explain our vivid percepts of motion. For this reason, Brentano and Husserl proposed that consciousness is restricted to a moment of time but that these moments contain true motion information. As Dainton pointed out, each time point on the physical clock time axis contains a second axis of mental time with exactly the information that is perceived during the physical moment.[3] Of course, one may ask what makes this second axis different from memory. Finally, extensionalism proposes that consciousness lasts for a substantial amount of time rather than being restricted to short moments. Often, these conscious periods are identified with the specious present, which is experimentally estimated to vary between a few hundreds of milliseconds to 3 seconds.[4] In all of these philosophical models, we are always conscious, which may not be surprising since philosophy usually relies on introspection, making it impossible to realize »gaps« in consciousness. Hence, the main question in the philosophy of the temporal aspects of consciousness is concerned with the duration of the chunks of consciousness (infinitesimally short or substantial, e.g., longer than half a second).

In psychology, for more than a century, the discussion has rather focused on whether consciousness is discrete instead of continuous, i.e., restricted to certain moments in time. During a discrete moment, by definition, no temporal information is available. Discrete perception is motivated by examples known at least since ancient Buddhist times. Consider the following sentence: »the mouse was next to the computer«. Reading this sentence, did you think about the cheese-loving animal or a computer mouse? Often, people think only about the latter because the later coming word »computer« disambiguates the former coming word »mouse«, i.e., well before a conscious percept was elicited. In a strictly continuous model, a thought about the animal would consciously be evoked before the word »computer« is heard.

If we combine the philosophical and psychological models, we arrive at 6 models (see Figure 1).

3 See Dainton 2018. These models are called retentional models. See Figure 1.
4 See Pöppel 1997 and Figure 1.

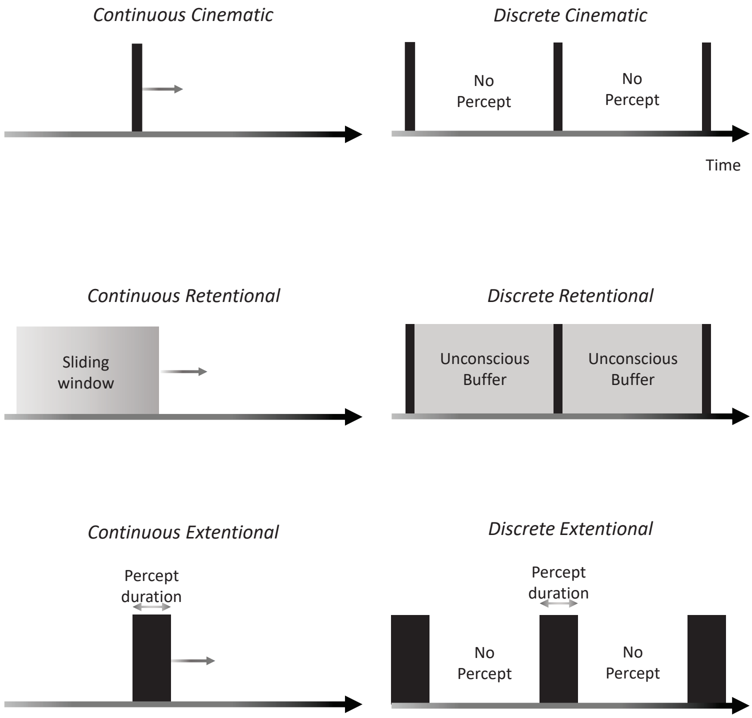

Figure 1.

Continuous models are presented on the left, discrete models on the right. The vertical bars indicate the duration of consciousness, which is (infinitesimally) small in cinematic and retentional models and extended in extensional models. In discrete models, conscious events are separated by periods of no consciousness. The horizontal line in the continuous retentional model indicates the time axis with the temporal information available at this moment in time.[5]

Effects similar to the mouse example, but experimentally better controlled, have been well known in vision research for a long time, such as apparent motion (see Figure 2). These temporal illusions have triggered an intense debate in both neuroscience and philosophy, with very little crosstalk between the two fields. In philosophy, the surplus

5 Adapted from Herzog et al. 2020.

content argument states that retentional models predict too many percepts. In apparent motion, for instance, one should first see a static disk, then motion, and finally the second static disk (see Figure 2). However, this is not the case. We *only* see motion. More complex models, including low-pass filtering, do not do any better.[6] In neuroscience, it is hotly debated when a conscious percept is elicited. Brain time apologists propose that we perceive an object as soon as its representation is elicited in the brain. In other words, time is presented by time. Event time apologists, to the contrary, propose that perception can be delayed in order to account, for example, for latencies between audition and vision. Since auditory processing is faster than visual processing, one would perceive the sound of a spoken word even before one sees the lips moving, similar to a thunder preceding lightning. However, this is not the case. The debate in both philosophy and neuroscience centers on effects and illusions based on delays shorter than 100ms and focuses on the question whether neural delays in the brain can account for these effects.[7]

Recently, a series of postdictive effects have changed the discussion. In these effects, for example, features of elements have been shown to be mandatorily integrated for several hundreds of milliseconds without conscious access to the single elements. As in the example of apparent motion, consciousness cannot have occurred before the last element was presented; thus, consciousness is substantially delayed. These delays are too long to be explained in terms of simple neural latencies. Figure 3 shows the sequential metacontrast paradigm (SQM).[8]

6 See Herzog et al. 2020.
7 See Holcombe 2013, Arstila 2015 and Piper 2019.
8 See Drissi-Daoudi et al. 2019.

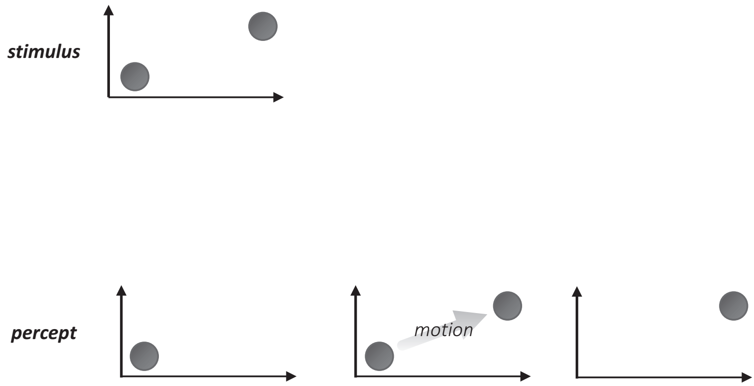

Figure 2.

Apparent motion. One disk is presented for a short duration, followed by a blank interval, and a second disk at a nearby location. We perceive motion from the first to the second disk, even though both disks are static. Importantly, we perceive the motion before the second disk is presented. However, motion direction can only be known after the second disk has appeared. A »too many percepts«-problem occurs for continuous models. In continuous models, one should see the first static disk during the time when it is presented (after some short neural delay), followed by a blank screen and afterwards the third static disk. Finally, one may perceive the illusory motion. However, humans have only one percept and not many.

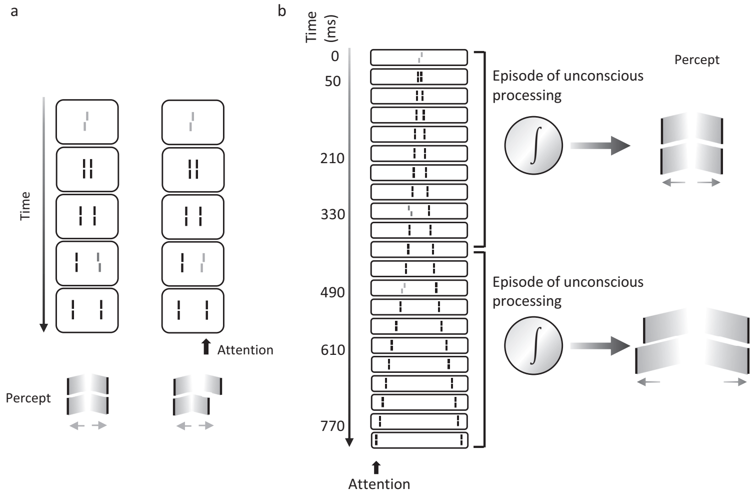

Figure 3.

An example of a long-lasting postdictive effect, the sequential metacontrast paradigm (SQM). A line with a horizontal offset, a so called vernier, is presented on a computer screen. The line is offset either to the left or right, randomly chosen. If presented alone, the vernier and its offset are clearly visible (not shown). **a)** When the vernier is followed by a sequence of lines without offset, the vernier is rendered unconscious. One perceives a stream of lines or, equivalently, two lines expanding from the center. In an experiment, the sequence of lines was presented either with or without the first vernier (the without offset condition is not shown here). Observers could not tell whether or not the vernier was present.[9] This effect is called metacontrast masking and is known for a century. Next, observers are asked to attend to the left stream of lines and report the offset direction of the visible lines. Even though the lines are all not offset and the vernier is invisible, observers reported the offset direction of the unconscious vernier. Hence, even though invisible, the offset survives metacontrast masking and is integrated with the visible lines. When one of the lines is offset itself, the two offsets integrate when the offsets are separated by less than approximately 400ms (right streams). If vernier and line are offset into the same direction, the percept of the offset becomes stronger (right stream of right example). When they are offset in opposite directions (right stream, left example), no offset is perceived on average. Importantly, integration is mandatory, i.e.,

[9] Cf. Otto 2006.

observers cannot report the first or second offset separately, even when asked explicitly. If, however, the second offset is presented later than 500ms after than the first one, both offsets can be reported (not shown). They perceive only one integrated offset. **b)** We used streams with more lines and offsets. In this example, the vernier offset and the opposite offset at 330ms integrated mandatorily but the offset at 490ms could be reported independently of the previous two, as we like to argue, because it is in a different period of discrete processing. Importantly, the first two offsets are 330ms apart from each other, the second and third offset only 160ms, hence, integration is not determined by temporal proximity *per se*.[10]

Simple neural delays cannot account for these results. Importantly, because these periods are so long, continuous models face a »too many percepts«-problem, which is similar to the surplus content argument.[11] In the SQM, for example, one would perceive the first offset consciously, then a series of straight lines, and finally the second offset. In continuous models, it is unclear when the integrated percept should occur or why integration occurs at all. Definitely, consciousness cannot occur before the onset of the second vernier (plus the basic neural latencies). However, as shown experimentally, observers have no conscious access to the individual verniers offsets (see Figure 3).

These results seem to favor discrete models, in which consciousness can be substantially delayed. However, traditional discrete models have considerable problems. For example, when we are conscious only at certain moments, it seems we would miss all the information in between the discrete moments because, by definition, a moment has no time. Moreover, how can we perceive motion at all when consciousness is short? Sometimes, it is proposed that motion occurs as change between the conscious moments. However, in this case, the discrete moments would need to be shorter than 3ms because we can perceive apparent motion on this time scale.[12]

10 Adapted from Herzog et al. 2020.
11 See Dainton 2018 and Grush 2016.
12 See Westheimer 1977.

2. The Many Components of the Problem

Physical Time

Thinking about time has changed physics strongly. Pre-Einsteinian physics has largely relied on a continuous flow of absolute time, which holds true for the entire universe[13] usually indicated with the letter *t* on the x-axis. This has changed with the theory of relativity where time depends on a reference frame. First, hence, there is no absolute time. Second, in the space time of general theory of relativity, the past and the future exist just as the present. Existence is not restricted to a physical moment. Third, an open question is whether physical time is continuous with infinitesimal small moments of time, well represented by the real numbers. It may well be that time is also discrete. In any case, time comes in quanta of a length of about 10^{-44}s, hence, much shorter than any neural time and hence too »fast«, as we would like to argue, for (conscious) perception.

Hence, our intuition about physical time, derived from our intuition about mental time, is not supported by physics. St. Augustine's conscious moments are not linked to an »objective« physical time. Conscious moments are not linked to physical moments. Hence, there is no gold standard of time we could use. As we will see next, it is also unlikely that mental time is directly linked to neural time.

Neural Time

The timing of neural activity depends on the type of neurons. Classically, neurons can elicit an action potential within 1–2ms, followed by a refractory period of a few milliseconds. However, since perceptual and behavioral measures tap into the response of the entire system or at least several thousands of neurons, temporal resolution may be even higher. Hence, in principle, neural resolution of humans could be in the millisecond regime, which is much lower than most measures of perceptual resolution.

Moreover, perception cannot be immediate because of neural latencies, e.g., from the retina to the brain as well as within the brain. These latencies depend on the contrast of the objects and are roughly

13 Galileo has already had the notion of a relative time, however.

in the range of 40ms to 100ms from retina to early visual areas of the brain. These delays are inevitable. However, visual processing takes often much longer before a conscious percept is elicited.

Mental and Temporal Resolution

First, we need to distinguish between the (specious) present, temporal resolution, perception of time, and the timing of a conscious percept. These aspects, even though interwoven, are all logically independent.

Since we cannot link the duration of the moments of consciousness to physical and neural time, we may link them to the perceived present, which is often called the specious present. This time is often estimated to be 2–3s long.[14] During this interval, we feel »now«. Empirical studies have shown that the estimation of the duration of time intervals is rather precise during this time. However, 2–3s are far too long for the minimal duration of consciousness because we can have numerous conscious events during this time.

A classic way to estimate the duration of consciousness is by assessing an observer's temporal resolution. For example, two bars are presented one after the other and observers are asked about their perception. For long durations, the two bars are perceived one after the other. When durations are short (e.g., 40ms), the bars are perceived as simultaneous, as if they were presented during the same moment of time. This makes it seem as if 40ms would be a good estimate for the duration of consciousness. One problem, however, is that different paradigms have led to very different estimates, ranging from 40ms to far beyond 100ms. An even more fundamental problem is that measures of temporal resolution are not necessarily measures of the duration of consciousness. A 40ms duration of simultaneity may simply reflect the temporal resolution of a simultaneity detector. For different tasks, such detectors have different resolutions. These measures are about the content of consciousness but not about the timing of consciousness *per se*. These two temporal aspects are logically independent.

It thus seems that we cannot measure the time of conscious perception directly. However, long lasting postdictive effects, such as the SQM, can provide upper and lower bounds. Since offsets are manda-

14 See Pöppel 1997.

torily integrated for about 400ms, this sets a lower bound. On the other hand, offsets beyond 490ms can be reported separately, thus setting an upper bound.

3. A Model

The provided examples of postdictive effects (computer mouse, SQM) show that conscious perception of a stimulus can occur much later after stimulus onset than expected from the basic neural delays. In particular, the relative timing of elements and their features is puzzling: we see a continuous stream of lines in the SQM. However, the combined offset we perceive enters into consciousness much later. Hence, the following question arises: what is perceived in the period before a conscious percept occurs?

We suggest: nothing. Conscious percepts are preceded by extended periods of unconscious processing during which we have no percept at all, i.e., consciousness occurs only at discrete moments of time. Each moment contains a »summary« of the entire event structure that happened during the unconscious processing. In this respect, our model is a discrete, retentional model with extended periods of unconscious processing, i.e., the second model in the second row of Figure 1.[15]

This model combines the advantages of both continuous and discrete models. Quasi-continuous, unconscious processing assures that information is continuously picked up from the environment and continuously processed. This raw input, however, is rather meaningless: For example, when an element is presented for the first time, it is unclear whether this element is a single static object that will stay confined to one location or will move in the next moment. Likewise, in the mouse example, the presentation of a first word leaves open many interpretations that depend on the subsequent words. We propose that humans have delayed consciousness to render only one interpretation of the stimulus visible. In this respect, slow consciousness, which contains a »summary« of the fast unconscious processing, makes sense. One may wonder whether unconscious processing has indeed processing capacities exceeding the ones of conscious processing by far. However, it is well established that unconscious processing

---

15 See Herzog et al. 2016 and Herzog et al. 2020.

is sufficient for carrying out complex cognitive tasks, such as understanding language or chess board situations. Likewise, as the SQM shows, fine-grained spatial and temporal features can be processed unconsciously with high capacity. Consciousness seems to come into play only when information is integrated across large parts of the visual field and semantic information becomes crucial, for example, when comparing two objects or searching for a person in a crowd.[16]

There are counter-intuitive aspects of our model. First, we intuitively feel that we perceive motion at each moment of time, as in the example of the diver above. The claim of the model is that we do not perceive the motion of the diver when it occurs. Instead, we perceive it much later and condensed. The situation is similar to the detection of other visual features. For instance, color is detected by dedicated color detectors, which for example indicate the saturation of a color by a number. Likewise, a higher luminance is coded by a larger number than a lower luminance, and a motion detector indicates how strong motion is by a number. In some way, the motion detector can be seen as mini-model of our model.[17] For example, when a line is moving, the line is first detected at one location and then at a second location. Because of a delay in neural latency (see Figure 4), the neural signals to both elements may arrive at the same time at a downstream neuron, which in this case elicits an action potential. If the arrival of the action potentials of the two neurons is not simultaneous, there is no action potential in the downs stream neuron M. A bank of neurons with different delays can thus deliver a set of detectors, which code for different velocities and motion directions. Importantly for our considerations, a neural response is only elicited after the second element is presented. Hence, the neural delay functions as a short-term memory. Such motion detectors can also explain why we are sensitive to apparent motion or, to turn it this way, why we perceive motion where there is none. For such detectors, it does not matter whether or not there is motion between the detectors (see Figure 4). In analogy to the motion detector, we can code entire event structures in a similar way. For example, when a line is presented for 30ms and followed by a square that is moving rightwards for 100ms, we perceive the entire event not during this epoch and not for 130ms but only for a short moment and much later, since it is coded by a corresponding event

16 See Mudri et al. 2014.

17 See Lee 2014 and Figure 4.

detector activated far after the event. Likewise, in the mouse example, information could be held in an unconscious memory before the entire information is rendered conscious.

Such event detectors can settle the memory vs. perception debate between cinematic and retentional models. For example, a motion detector makes up a retention without the necessity to propose a second time axis. As mentioned, there is memory through the neural latency in the detector. However, this memory is not consciously accessible and, thus, only a strong motion percept is elicited. Inferences, to the contrary, work by explicit memories where both elements are consciously perceived as in the example of the trash bin above.

In philosophical terms, the above considerations support the view that external world timing is not represented by neural time, much like color is not coded by colored neurons.

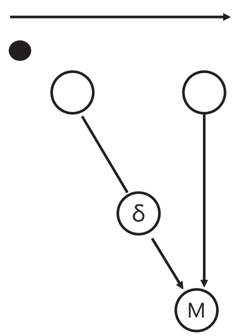

Figure 4.

A basic motion detector. When the dot moves from left to right, it first activates the left-most upper detector, which sends a signal to the downstream neuron M. The signal is delayed by a time . Then, the dot activates the right detector, which also sends a signal to M. When the two signals arrive at the same time at M, M also sends a signal indicating that there is left-right motion with a certain speed corresponding to .

4. Discussion

Dennett and Kingsbourne proposed that we cannot determine the microstructure of consciousness because we cannot discriminate between two scenarios: whether consciousness at one very moment is immediately suppressed (Stalinian scenario) or whether consciousness is retroactively modified (Orwellian scenario).[18] For example, in the SQM, one may well have a conscious percept of the first offset before the second offset is presented. This conscious percept does not enter memory and is suppressed when it is combined with the second offset. We cannot exclude the possibility of such fleeting consciousness. In this respect, our model only addresses consolidated consciousness, which is stable for at least a few seconds and could potentially be transferred into long-term memory.

One may argue that actions can be much faster than a few hundred milliseconds. While this is true, we argue that eliciting an action does not need consciousness as evident in reflexes and fast reactions in sports. We rather claim that actions are part of the event structures and are also perceived as such.

Further, one may argue that the existence of detectors coding for entire, complex events is highly unlikely, given the vast space of possible events. This may be true. Likely, neural coding is more complex than the outputs of simple detectors. However, this problem is not restricted to temporal events but holds true for static scenes as well. Humans have a strong capacity not only to distinguish many similar scenes, such as beaches, but can also memorize a large number of them. Thus, here is a general problem to solve.

One may also argue that the typical measure for neural responses are spikes per second, i.e., a measure extended in time. This is true. However, the timing of the spikes does not need to go hand in hand with the timing of the temporal events. However, as it is largely unknown what the neural code is, new insights may lead to different interpretations.

If perception is discrete, why do we not realize the gaps in between conscious events? This is simply because consciousness is, by definition, slower than the gaps. A system cannot go beyond its own resolution or, to turn it this way, we can only be conscious in and about the conscious events when we are conscious and not when we are unconscious. For this reason, conscious events follow each other smoothly in conscious time but not in physical time, which is consciously not accessible but likely unconsciously. In this respect, conscious time is shorter than physical time. However, we may still feel

18 See Dennett & Kingsbourne 1992.

an extended consciousness because unconscious duration detectors may determine exactly the time elapsed during unconscious processing and render this number conscious together with all the other information.

The surplus content argument was originally used by Dainton to argue against retentional models (Grush, 2016). However, as shown with the examples of the SQM and apparent motion, the surplus content argument argues not against retentions but against continuous consciousness including extensionalistic ones. Our discrete model avoids this problem because there is not a percept at each moment of time. If there were only two conscious percepts per second, there would definitely not too many percepts. Traditional discrete models have a similar problem because they are usually (implicitly) continuous: we are conscious at each moment of time, but percepts are constant during extended moments because a moment cannot contain any temporal information. We avoid this problem for the same reason as above: there are no continuous percepts and hence no constant percepts. The unconscious processing periods therefore solve the problems of both approaches in the same way.

There is a high price to pay: the model is at odds with how we feel our perception comes about. For example, in some instances, we argued that certain percepts are simply not what we believe they are, whereas, in other instances, we heavily relied on a certain percept being true. For example, we believe we see a continuous stream in the SQM, but our model claims this is an illusion. On the other hand, we claimed that there is no conscious percept of the first offset. Hence, one may prioritize the percepts this way or the other way around and arrive at different conclusions. An additional problem is that phenomenology becomes unreliable when we approach the temporal limits of conscious time. It is usually difficult to know what we perceive when time is short. This seems to be a general problem of the timing of consciousness. The future will show whether there are models that can marry our feelings about phenomenology with explanations for the puzzling results from temporal illusions.

Acknowledgments

I like to thank Lukas Vogelsang and Maëlan Menétrey for comments on the manuscript. This work was supported by the Swiss National

Science Foundation grant ›Basics of visual processing: from elements to figures‹ (176153).

References

Arstila, Valtteri (2015): Defense of the brain time view, in *Front. Psychol.* 6.
Dainton, Barry (2018): Temporal consciousnes, in *The Stanford Encyclopedia of Philosophy*.
Dennett, Daniel & Kinsbourne, Marcel (1992): Time and the observer: the where and when of consciousness in the brain, in *Behav. Brain Sci.* 15, 183–201.
Drissi-Daoudi, Leila, Doerig, Adrien, & Herzog, Michael (2019): Feature integration within discrete time windows, in *Nature Communications* 10(1), 1–8.
Efron, Robert (1967): The duration of the present, in *Ann. N. Y. Acad. Sci* 138, 713–729.
Grush, Rick (2016): On the temporal character of temporal experience, its scale non-¬-invariance, and its small scale structure, Manuscript, doi:10.21224/P4WC73
Herzog, Michael, Drissi-Daoudi, Leila & Doerig, Adrien (2020): All in Good Time: Long-Lasting Postdictive Effects Reveal Discrete Perception, in *TICS 24*, 826–837. https://doi.org/10.1016/j.tics.2020.07.001.
Herzog, Michael, Kammer, Thomas & Scharnowski, Frank (2016): Time Slices: What Is the Duration of a Percept, in *PLoS Biology* 14(4), doi:10.1371/journal.pbio.1002433.
Holcombe, Alex (2013): The temporal organization of perception, in Wagemans, J. (ed.), *Handbook of Perceptual Organization*, Oxford University Press, 820–840.
Lee, Geoffrey (2014): Temporal experience and the temporal structure of experienc, in *Philos. Impr*. 14, 1–21.
Mudrik, Liad et al. (2014): Information integration without awareness, in *Trends Cogn. Sci.* 18, 488–496.
Otto, Tom, Ögmen, Haluk & Herzog, Michael (2006): The flight path of the phoenix— The visible trace of invisible elements in human vision, in *Journal of Vision* 6, 1079–1086.
Piper, Matthew (2019): Neurodynamics of time consciousness: an extensionalist explanation of apparent motion and the specious present via reentrant oscillatory multiplexing, in *Conscious. Cogn.* 73, 102751.
Pöppel, Ernst (1997): A hierarchical model of temporal perception, in Trends Cogn. Sci. 1, 56–61.
Westheimer, Gerald & McKee, Susan (1977): Perception of temporal order in adjacent visual stimuli, in *Vis. Res.* 17, 887–892.