**Retrodictive and Predictive Attentional Modulation in Temporal Binding**

Rasmus Pedersen

Department of Philosophy

University of Sydney

**Affiliation:** University of Sydney, Department of Philosophy, Sydney, Camperdown NSW 2050, Australia

**Email:** rped0728@uni.sydney.edu.au

**ORCID iD:** 0000-0003-2261-0582

# Abstract: This paper sets forward a novel theory of temporal binding, a mechanism that integrates the temporal properties of sensory features into coherent perceptual experiences. Specifying a theory of temporal binding remains a widespread problem. The popular ‘brain time theory’ suggests that the temporal content of perceptual experiences is determined by when sensory features complete processing. However, this theory struggles to explain how perceptual experiences can accurately reflect the relative timing of sensory features processed at discrepant times. In contrast, ‘event time theories’ suggest that the temporal content of perceptual experiences reflects the relative event time of external sensory features and that the brain accommodates differential processing times. We can formulate retrodictive and predictive versions of this theory. Retrodictive event time theories propose that we accommodate desynchronised processing retrodictively. Predictive event time theories propose that we accommodate desynchronised processing by predictively modelling the event time of sensory features. I argue that both views have strengths and weaknesses. This paper proposes a new hybrid theory that integrates these theories to accommodate these weaknesses. Firstly, I argue how retrodictions and predictions can interact in mutually beneficial ways to ensure speedy and accurate temporal binding. Secondly, I propose how attention plays a central role in flexibly selecting which contents get to be temporally bound. This theory can explain how prediction and retrodiction differentially affect temporal binding and, in turn, proposes a new way to understand temporal binding and has implications for how we should understand conscious experiences.

**Keywords:** Temporal Binding · Perceptual Experience · Predictive and retrodictive models · Attention · Event time theories

# 1 Introduction

The temporal binding problem is the problem of how temporal properties of sensory information are integrated into coherent representational states so that we can consciously perceive the temporal order, synchrony and change of external events. Temporal binding is complex because sensory features that occur at the same event time propagate at different speeds hit our sensory receptors at different times, have discrepant processing times, and can influence each other’s processing speed (Bullier, 2001; Groen et al., 2013; Pöppel, 1988; Zeki, 2015). Given all these potential desynchronisations, how can we perceptually reflect what happens when in ways that allow us to interact competently with the external world? The functions and neural basis of temporal binding denote a widespread problem in psychology (Nishida & Johnston, 2010). In this paper, I attempt to provide a new functional theory of how our brain engages in temporal binding.

One popular kind of theory of temporal binding called the brain time theory states that the sensory contents that get temporally bound and, as such, are perceptually experienced as synchronous are determined by the time at which those sensory contents finish processing and reach the process(es) responsible for conscious perceptual experience (for supporters see (Arnold et al., 2001; Arstila, 2015a, 2016a; Moutoussis, 2012; Piper, 2019; Yarrow & Arnold, 2016; Zeki, 2007; Zeki & Bartels, 2009). A general commitment of this kind of view is that the time represented (the temporal content represented in perceptual experiences) is isomorphic to the time of representing (the time at which the underlying mechanism represents the content of the perceptual experience). I hold that, due to this commitment, the brain time theory cannot account for a range of ways in which temporal binding and, thus, the temporal content of our perceptual experiences (time represented) is affected without these effects being the result of changes in neural processing times or necessary processing delays (the time represented).

Psychophysical evidence of predictive and retrodictive effects on temporal binding indicates that there is a trade-off between how quickly temporally bound perceptions are made cognitively accessible (speed of perceptual availability) and how accurately they represent the event time relations of external sensory features to be bound (accuracy of temporal binding). The complexity of the stimuli, the behavioural task one is engaged in, and how one’s attention is directed influence the speed and accuracy of temporal binding. A theory of temporal binding must explain how the brain engages in this trade-off in ways that account for the relevant psychophysical and neurological data, and this involves abandoning the idea of the brain-time theory that the time represented and the time of representing cannot come apart.

One can do this by embracing an event time theory. According to event time theories, temporal binding is aimed at binding temporal properties of sensory features relative to the time at which they occur in the external world and not relative to when they finish processing in the brain, as per the brain time theory (Nishida & Johnston, 2002).

I take it that there are two general ways to spell out an event time theory that breaks the relative ordinality between neural processing times (and thus the time of representing) and the time represented. On one side, there are existing retrodictive event time theories that argue that this ordinality is broken retrodictively (Dennett, 1991; Dennett & Kinsbourne, 1992; Johnston & Nishida, 2001; Libet et al., 1979; Nishida & Johnston, 2002). On the other side, there are a range of authors suggesting how this ordinality can be broken predictively, which I argue can be used to describe a predictive version of the event time theory (Bechlivanidis et al., 2022; Bechlivanidis & Lagnado, 2009; Blom et al., 2020; Demarchi et al., 2019; Hogendoorn, 2022; Hubbard, 2014; Kwon et al., 2015; Nijhawan, 2008; Toren et al., 2020; White, 2018). I will call these retrodictive event time theories and predictive event time theories.

I hold that both these kinds of theories face some issues. Retrodictive, even time theories generally fail to account for how we manage to engage in temporal binding at speeds that allow us to perceive certain events as occurring in actual event time. Predictive event time theories only explain temporal binding in a predictive context and fail to account for how our perceptual experiences reflect more or less accurate temporal bound representations in unpredictable situations.

In this paper, I propose a hybrid theory of how our brain makes use of a combination of retrodictive and predictive processing to flexibly trade between speed and accuracy when engaged in temporal binding. I do so by proposing two integrated hypotheses.

The first integrated hypothesis proposes that retrodictive processing and predictive processing interact in mutually beneficial ways to temporally bind sensory information. Predictive processing allows for speedy temporal binding, and retrodictive processing allows for the redrafting of inaccurate predictive perceptions as processing deepens over time. By integrating these, we can solve the respective problems of predictive and retrodictive processing. This integrated hypothesis can be seen as an extension of Grush’s (2005) and Hogendoorn’s (2022) existing hybrid theories of trajectory estimation.

The second integrated hypothesis concerns how these two kinds of processing relate to the content experienced in conscious experiences. I present a model of how attention plays a central role in modulating which contents do and which do not get to be temporally bound into conscious perceptions. By explicating the selective role of attention, we can understand how temporal binding is sensitive to both top-down influences (causal inferences, behavioural tasks, perceptual goals) and bottom-up influences (environmental context, surprising/attention-grabbing features). As such, this theory also gives rise to a range of empirical questions/predictions and allows us to explain how prediction and retrodiction differentially affect temporal binding.

In section 2, I set out the notion of temporal binding and the problems related to accounting for temporal binding. I also provide a gloss over why the brain time theory fails to account for these problems. In section 3, I provide evidence for how temporal binding is subject to predictive and retrodictive effects in ways that seem to make temporal binding affected by goals, environmental context, perceptual tasks, expectations and attentional probing. In section 4, I overview some of the more popular theoretical frameworks for thinking about how the brain exploits retrodictive processing for temporal binding and how it may exploit predictive processing for temporal binding. I point out the shortcomings of both kinds of frameworks in explaining the data set out in section 3. This motivates my hybrid, which promises to solve these problems by providing a single framework for how retrodictive and predictive processing interact in mutually beneficial ways that allow for it to explain the flexibility in temporal binding that evidence seems to demonstrate. In section 5, I set out the two integrated hypotheses of this hybrid view and how it can explain the data set out in section 3. In section 6, I present some general consequences of this view and some empirical evidence that uniquely supports my hybrid theory.

# 2 Dealing with a Trade-off for Temporal Binding

I take there to be a selective trade-off involved in temporal bound perception between ‘the speed of perceptual availability’ and ‘the accuracy of temporal binding’. The first is about the speed at which sensory features are made consciously and cognitively accessible for perceptual purposes. The second refers to how accurately a subjective percept represents the relative times at which features occur in event time.

As such, there is not only selection pressure for mechanisms that make present events perceptible in time (minimising delay) but also selection pressure for mechanisms that compensate for differential delays to allow unconscious preparation of a percept (Changizi, 2008).

It may seem optimal for an organism to minimise delays as much as possible, for example, by favouring short neural pathways and fast transduction mechanisms. However, I suggest another picture. Allowing for more processing time increases the possibility for recurrent processing, which in turn can deepen cortical processing and the details of a percept (O’Reilly et al., 2013). Longer processing times do not just mean longer delays but also more accurate perceptions. Neural delays aren’t only a biological processing constraint that cannot be fully overcome. Allowing for delays also provides a way for the brain to buy time for preparing a perception. As such, there would also be selection pressure for compensation mechanisms that makes it possible for such a time-consuming deepening of perceptual processing without it affecting our ability to accurately perceive the time of external events.

Delaying perceptual availability allows the brain to temporally bind external features into the same percept accurately. If sensory information is made perceptually available the moment features are processed, then chances are that we fail to perceive sensory features occurring together in event time in the same percept. If, on the other hand, our system emphasises accurate temporal binding over all else, this comes at the cost of how quickly perceptions are consciously accessible – which would make it impossible to perceive events in real time and act accordingly.

As such, I emphasise that in the case of temporal binding, there is a trade-off to utilise time wisely to make perceptions available in time but allow for enough delays to properly prepare perceptions.

This is in contrast to the claims of defenders of the brain time theory (Arnold et al., 2001; Arstila, 2015, 2016a, 2016b, 2019; Engel & Singer, 2001; Moutoussis, 2012; Piper, 2019; Yarrow & Arnold, 2016; Zeki, 2007, 2015; Zeki & Bartels, 2009). They claim that temporal binding is determined by processing times – so that synchronously processed features are perceptually experienced as synchronous. If this is true, then the speed at which sensory features finish processing and reach the relevant process for conscious experience fully determines the accuracy of temporal binding. But this seems to be false because it cannot account for psychophysical data that show how the temporal content of our perceptual experiences can be influenced without any influence in event-related brain potentials, i.e., neural processing times. Below, in section 3, I describe a range of cases where the temporal content of our perceptual experiences is affected by retrodiction and prediction in ways that cannot easily be explained by appeal to differential neural delays. For these reasons, it seems that the brain time theory is getting the functions of temporal binding wrong in so far as we cannot rely only on temporal processing times, differential neuronal latencies, and the temporal resolution of different local and parallel processing of sensory features if we want to account for how our brain engages in temporal binding.

The cases presented in section 3 bolster the claim that our perceptual system makes use of retrodictive and retrodictive processing when engaging in temporal binding I provide a range of concrete examples of how our perceptual system retrodictively and predictively breaks the ordinality between neural processing times of stimuli and the temporal properties attributed to those stimuli in perceptual experience. I do this to motivate the move away from a brain time theory and towards an event time theory, according to which this ordinality can be broken to flexibly trade between speed and or accuracy.

# 3 Psychophysical Evidence for Retrodiction and Prediction in Temporal Binding

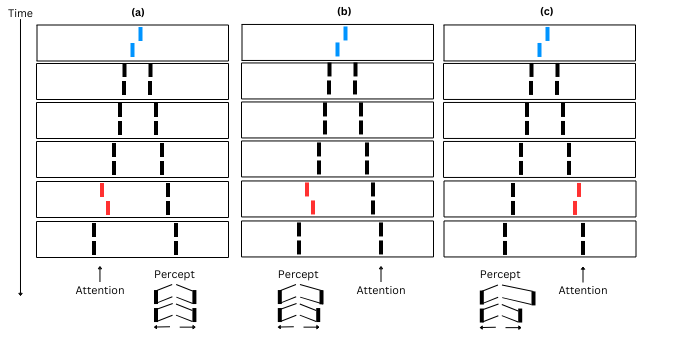
In this section, I present a range of studies supporting that temporal binding can be affected retrodictively and a range of studies of how temporal binding can be affected predictively. In section 4, I then provide retrodictive event time theories that can account for the retrodictive effects and predictive theories that can be used to formulate a predictive event time theory that can account for the predictive effects. This naturally motivates my proposal of a hybrid theory in section 5 that integrates these theoretical frameworks to explain the data presented in section 3.

## 3.1 Evidence for Retrodiction in Temporal Binding

Retrodictive (postdictive) effects describe visual phenomena where the perception of a stimulus occurring at event time t2 affects the perceptual experience of a stimulus occurring at event time t1. Typical examples of this include cases such as the cutaneous rabbit, where evenly spaced taps on the arm are experienced as gradually moving up the arm (Geldard & Sherrick, 1972). Various versions of the colour-phi phenomenon, where two dots flashed in succession at two different locations and in two different colours, are experienced as a single continuously moving dot that changes colour midway (Kolers & von Grünau, 1976; Ramachandran & Anstis, 1983). Cases of backwards masking, where a stimulus occurring presented at t2 masks the perception of an earlier stimulus presented at t1 (Shimojo, 2014). These classic examples have been used to argue for the role of retrodictive processing in temporal binding and for the fact that the time represented can come apart from the time of representing (Dennett & Kinsbourne, 1992; Eagleman & Sejnowski, 2000; Grush, 2005; Hogendoorn, 2022; Lee, 2014). While they are promising, the timescale across which the retrodictive integration of stimuli occurs in these cases is often so short (less than 100ms) that it is difficult to know whether cases of retrodiction that cannot simply be explained by long neural integration times or as the result of recurrent sensory processing (see Arstila, 2015, 2016a, 2016b; Piper, 2019). If this is the case, then it is not clear at all that the time of representing and the time represented come apart. For example, Lamme & Roelfsema (2000) argue that the content of perceptual experiences is determined within 100 to 150ms following stimulation. If this is right, then retrodictive effects that fall below this threshold might not need to indicate any problem for brain time theories.

But there is also evidence for cases where the retrodiction takes place over a much longer period that cannot easily, if at all, be jotted down to long integration times or recurrent processing loops. Herzog et al. (2020) discuss a variety of cases of perceptual retrodiction that take place across several hundreds of milliseconds.

To get a grasp of how strongly retrodictive processing can influence the temporal binding of sensory signals, consider the following experiment by Drissi-Daoudi et al. (2019). They use a sequential metacontrast paradigm where two sets of stacked vertical lines are presented in sequence to prompt a percept of two diverging streams. When the horizontal lines of one of the sets have a minimally perceptible offset, all other lines in the sequence are perceived as offset in the same direction, even though they are straight (see Figure 1). This offset can be reported by the observer, yet it is impossible to tell which of the set of lines in the sequence is ’carrying’ the actual offset – the offset contaminates the rest of the lines in the sequence. If multiple minimally perceptible offsets are present in one of the streams, then the offsets are bound and cannot be reported independently. If these offsets are present in different streams, they are not bound, so they are stream-specific – probably because they are separately attended to. If two offsets are shown in the same stream but in opposite directions, any offset in the entire stream is cancelled out, if the offsets are in the same direction, the perceived offset of the entire stream is increased. This temporal binding of offset lasts up to 450ms (Drissi-Daoudi et al., 2019). But our conscious perceptions of the offset line segments are not delayed several hundreds of milliseconds; rather, they are more or less immediately available, it is just that our perceptions seem to remain retrospectively editable. In section 3.1, we explore theories of how such retrospective edits may take place.



**Figure 1:** (a) when attention is allocated to a stream where two offsets are present in the opposite direction the perceptual discrimination of any offset cancels out. (b) when attention is allocated to a stream where one offsets is present the offset contaminates the entire stream. (c) when two offsets in the same direction in the same stream are present it increases the offset perceived across the entire stream.

Other studies exhibit a similar effect across modalities. In an experiment by Stiles et al. (2018), three flashes were presented in sequence simultaneously with a sound so that these seemed paired. This has the effect that when the flash in the middle of the sequence was omitted, a central flash was illusorily perceived where only the sound was played. Conversely, when the middle was omitted, but the flash was present, the flash was illusorily perceived as absent. This kind of experiment exhibits how our perceptions of auditory and visual stimuli “can be illusorily suppressed or added postdictively across hundreds of milliseconds” (Herzog et al., 2020).

Moreover, as I will also emphasise later, Rimsky-Robert et al., (2019), Sergent et al., (2013), and Thibault et al., (2016) show how cueing attention to a stimulus retrospectively, multiple hundreds of milliseconds after stimulus offset, can influence our perceptual experience of that stimulus. By drawing attention to the location of an earlier otherwise near-threshold (effectively imperceptible) stimulus, the chances of experiencing that stimulus as occurring at that time increase.

These long-term cases of retrodiction provide evidence for how retrodiction can influence temporal binding without necessarily affecting the speed at which sensory features are processed. It is hard, if not impossible, for a brain time theory to account for these retrodictive effects. If the brain time theory is on the right track, then neural processing of the features that are temporally bound bounces around in our perceptual system for up to 450ms so that the later occurring stimuli can affect earlier occurring stimuli by reaching the processes responsible for conscious experience simultaneously. This is an extremely unlikely explanation since it would mean that the necessary delay involved in processing this kind of stimuli is almost half a second. This problem indicates a more general problem with the brain time theory, which is that it does not explain how we can perceptually experience two features with discrepant processing times occurring together.

As we will discuss below, retrodiction aids temporal binding in the way that when we are faced with an event composed of an array of different features that are processed at discrepant times, such discrepancies can be retrospectively ironed out so that all features of one event are temporally bound into a single coherent perception. As such, retrospective processing allows us to explain how temporally bind the features of complex events without our perceptions being much affected by differential neural latencies.

## 3.2 Evidence for Prediction in Temporal Binding

In the last decade, there has been a large emphasis on predictive aspects of conscious perception (Clark, 2013; de Lange et al., 2018; den Ouden et al., 2012; Dolęga & Dewhurst, 2020; Friston, 2010; Hohwy, 2012, 2013; Knill & Pouget, 2004; Nijhawan, 2008; Parr & Friston, 2017; Rao & Ballard, 1999; Summerfield & de Lange, 2014; Walsh et al., 2020; Wiese & Metzinger, 2017). In this section, I want to present some of the evidence in support of the fact that prediction influences temporal binding. Explicitly spelling this relation between temporal binding and prediction is becoming increasingly important as a large range of authors implies that prediction has implications for time perception (Bechlivanidis et al., 2022; Bechlivanidis & Lagnado, 2009; Blom et al., 2020, 2021; Demarchi et al., 2019; Hogendoorn, 2022; Johnson et al., 2023; Kwon et al., 2015; Nijhawan, 2008; Pariyadath & Eagleman, 2007; Toren et al., 2020; Turner et al., 2024; White, 2018).

In the case of temporal binding, prediction is useful to overcome the issue of neural delays involved in processing as well as the problem of discrepant processing times. Firstly, if external sensory features’ temporal properties can be predicted before they are directly processed, then the time of representing can take place closer to the event time and well in advance of direct processing. Secondly, if the temporal properties can be predictively temporally bound with each other, then the time represented can reflect the relative event time of sensory features despite differential processing times. So, if there are predictive effects on temporal binding, then direct processing times do not determine the time represented nor the time of representing as claimed by the brain time theory.

In this section, I present two ways in which I take prediction to play this role in temporal binding. These are 1. feedforward predictions and 2. feedback predictions.

*1. Feedforward prediction:* The brain makes use of feedforward processing and simple memory functions to extrapolate a temporal estimate of when features of a set of events occur relative to each other – these are all mechanisms of an evolutionary or developmental origin (Rust & Palmer, 2021).

*1a Representational momentum:* One example of feedforward prediction is motion extrapolation, such as the kind involved in representational momentum (Hubbard, 2014; Thornton & Hubbard, 2002). Representational momentum involves watching a sequence of stimuli (or a video) of a rotating or moving object shown and then subsequently judging which probe stimuli is identical to the last frame of the presented sequence. The common finding in such cases is that participants tend to choose (at the above chance) a probe that slightly overshoots the last frame of the sequence. They choose the ‘potential next frame’ implied by the momentum of the sequence (Thornton & Hubbard, 2002).

Kimura (2021) has reviewed representational momentum as a case of predictive processing. By predictively modelling the momentum and trusting our prediction, we fail to choose the right probe. Others have argued that representational momentum is not very well understood and could just as well be the result of memory bias (Halpern & Kelly, 1993). There are, however, several indicators that momentum-like effects are influenced by feedback about prediction errors (Ruppel et al., 2009), expectations about direction (Hubbard & Bharucha, 1988), and expertise in specialised momentum cases (Blättler et al., 2010). This suggests that people draw on past actions or states of a visual target to anticipate its future state, which supports that representational momentum is an automatic perceptual consequence of a prediction.

*1b Trajectory estimation:* A series of studies by Blom et al., (2020), Blom et al., (2021), Johnson et al., (2023), and Turner et al., (2024), indicates how prevalent predictions are when it comes to perceiving trajectories. Taken together, these studies suggest how extrapolation occurs during early visual (retinal) processing and Turner et al., (2024) show how extrapolation also occurs during a series of later stages of (cortical) visual processing. The extrapolation seems to cause visual perceptual representation to shift closer to event time, enabling us to perceive motion trajectories live as they unfold. This is a clear case of how predictive processing plays a direct role in our capacity to temporally bind visual features across different levels of processing, thus overcoming delays otherwise involved if our perception was based on direct processing with no extrapolation.

*1c Predictive updating:* There is also a range of cases where the perception of a repeated temporal offset of a pair of stimuli leads to the recalibration of the perceived timing of the offset (Di Luca et al., 2009; Fujisaki et al., 2004). Participants repeatedly presented with a sound and a flash 235ms apart will, within 3 minutes of exposure, shift their point of subjective simultaneity by approximately 22ms to compensate for the offset (see Fujisaki et al., 2004). This compensation shows up in multiple modality pairings (Di Luca et al., 2009). Moreover, these compensations do not seem to be caused by shifting the time of stimulus-evoked neural signals (Holcombe, 2015, 829). According to Holcombe (2015, 830), the rationale for a shift away from asynchrony toward synchrony stems from statistical assumptions about the natural environment, where “the distribution of the relative timing of stimulation by external events is likely to be centred on or near zero”. This would be encoded in a predictive model. This combination of statistical priors and the adaptation to the delay in the (initially novel) sequence can be straightforwardly explained as a learned prediction and shows how predictive updating influences how we temporally bind sensory features without any changes in the time of stimulus-evoked neural signals.

*2. Feedback predictions*: There is also evidence for predictive effects that stem from feedback processing from more complex memory structures—these rely on learning in adulthood rather than on an evolutionary and/or developmental basis (Rust & Palmer, 2021).

*2a Mnemonic expectations:* When trained to learn a sequence of images, actions or consequences in the form of visual images, people become very proficient in perceptually predicting what new stimuli are likely to occur in the near future (Hindy et al. 2016). Hindy et al. (2016) call these mnemonic expectations, which include predictively processing the identity of familiar things that we expect to see in our near future, e.g., when expecting to see a specific building when turning a corner in a familiar neighbourhood. Such mnemonic expectations not only activate the same areas as feedforward visual predictive mechanisms but also feedback activity, such as the retrieval of past experiences. They argue that this retrieval mechanism is a result of hippocampal pattern completion, where exposure to a familiar past event activates a ‘conjunctive representation’ of the entire experience. This results in an ability to prepare perception for complex temporal binding of sensory features based on higher-order feedback mechanisms. Other evidence for this includes cases where neural firing rates of higher stages of visual processing (inferotemporal cortex neurons) increase in firing rate when perceiving a violation of familiar sequences (Meyer et al., 2014).

*2b Predictive overpowering:* In a recent study, Bechlivanidis et al. (2022) show how our time perceptual system tends to be biased towards familiar causal relational patterns. In this study, participants view a causal sequence of A causing C causing B. Yet due to causal expectancies, this sequence is perceived as the sequence A-B-C thus reversing the actual sequence of CB to the perceived sequence BC. Participants are fully attending and allowed to view the sequence in repeat, yet the reversal effect persists. This reversal effect indicates that there are predictive that overpower the external visual signals. These results are easily interpretable if we take perception to utilise encoded top-down causal expectations to guide the perception of temporal order. This perceptual reversal of the causal sequence should be interpreted as the result of top-down feedback processing and complex memory functions because it is the persisting belief based on complex ingrained memories of causal structures that influence the perceptual extrapolation effect. The extrapolation is not due to the effect of short-term memory or some extrapolation done through continued feedforward processing of the stimuli. This is evident from how the reversal effect in Bechlivanidis et al. (2022) persists even when repeated.

There are many similar examples where the ordinality of brain time is broken in favour of highly ingrained top-down expectancies of systematic changes in the external world. We see this in cases of temporal ventriloquism (Hartcher-O’Brien & Alais, 2011; Vroomen & De Gelder, 2004), in situations where our brain seems to compensate for the relative differences in the propagation speed of sound and light (see Alais & Carlile, 2005), and where self-induced cause and effect relations influence the temporal order perceived (Haggard et al., 2002; Stetson et al., 2006) , which should also be treated as the result of top-down influences given that causal representations are necessary for the effect to take place (see Hoerl et al., 2020, 23; Tramacere & Allen, 2022, 14).

With these examples, we see how predictive processing influences temporal binding by displacing direct processing with feedforward and feedback perceptual predictive models of temporal relations. These predictions affect both the time of representing and the time represented. Predictions can shift the time of representing closer to event time this helps the perceptual system to represent features in due time despite neural delays. Moreover, predictions can influence which temporal properties are bound independently of the differential timings of direct neural processing, thus counting as evidence against the brain time theory.

# 4 Perceptual Theories of Retrodiction and Prediction

While there is a growing consensus that our time perception is influenced by retrodictions and predictions, there is no consensus about how we should go about explaining all these various effects. Below, In section 4.1, I introduce two event time frameworks that have taken time perception to utilise retrodictive processing by Dennett & Kinsbourne (1992) and Nishida & Johnston (2010). In section 4.2, I present a general theoretical framework on predictive processing that I propose can be neatly used to formulate an event time theory of temporal binding. I propose some benefits and shortcomings of these views and argue that these issues can be overcome with the hybrid theory that I propose in section 4.

## 4.1 Retrodicting Time

Retrodictive event time theories take it that the ordinality of when features are processed in specific mechanisms (content fixation) is retrodictively broken relative to the ordinality of our subjective experience (time represented) so that our temporally bound perceptions better represent the relative event time of external sensory features.

This can be accomplished by having a perceptual system with ‘time markers’ that timestamps sensory signals’ event time (by encoding signals’ sensory onset time). Any desynchronisation that might take place when these signals are processed in different mechanisms can be compensated for by retrodictively aligning the event time of processed signals with reference to encoded time markers. This kind of ‘time marker account’ has been proposed by Libet et al., (1979) and further developed by Nishida & Johnston (2001; 2010). Dennett and Kinsbourne (1992) posit a distinct system where any initial desynchronisation in the processing of sensory features is compensated for through temporal binding being sensitive to contents. The temporal binding of these contents is then done with regard to their ‘relative best phase of correspondence’, which is constrained by a range of top-down goals and bottom-up contexts. This is the ‘content-sensitive settling view’.

### 4.1.1 Nishida & Johnston time marker view

According to Nishida and Johnston’s time marker account (2010), the brain groups the temporal properties of sensory signals originating from a broad set of sensory mechanisms by having two distinct kinds of mechanisms in place that make it possible to encode time markers within a sensory channel and compare them across sensory channels. Nishida & Johnston posit early-level mechanisms that ‘mark’ the onset and transients of sensory signals. These are quick feature-specific mechanisms with a high temporal resolution. Because these early-level mechanisms work quickly and pick up on the onset of certain transients of sensory signals, they represent something akin or close to event time and do so in isolation of delays in later cortical processing of the features.

These ‘time markers’ can then be used as input to mid-level perceptual mechanisms that compare the temporal relations of features across modalities. The mid-level comparator is a comparatively slow, flexible, and amodal mechanism that “selects, individuates, and registers high-salience features” to make a “saliency-based cross-channel comparison” (Nishida & Johnston, 2010, 284). So, this mid-level mechanism maps the saliency of features and associates them with salient temporal features without carrying information about the features themselves. The time represented in perceptual experience is determined by the cross-channel association of time markers registered in early-level processing. This means that our perceptual experience can reflect the relative event time of sensory features despite later desynchronisation.

While Nishida & Johnston’s account may apply more widely, it is developed to provide an alternative theoretical explanation of the colour-motion asynchrony illusion. In this illusion, there is a pattern that oscillates between moving upwards and downwards, changing direction at some set rate of change. Moreover, the pattern changes colour as it changes direction, it is green when moving upward and red when moving upward. When the rate of change is high (1-2hz), people cannot confidently make the pairing between colour and direction and have difficulties with telling which direction goes with what colour. They fail to see the colour as changing simultaneously with the change of the direction—instead, they may perceive these features as changing asynchronously or just be unable to pair them. Yet if the time of colour change of the pattern is manipulated so that it is delayed about 100ms relative to the motion direction change of the pattern, then any problems in binding the two features disappear, and participants report that they perceive the features as changing occur simultaneously (Moutoussis & Zeki, 1997). This has been taken to support the view that it is differential processing times that explain the apparent asynchrony.

On the time-marker account, however, it is not differences in feature processing times but differences in the kinds of temporal structures of the features in the illusion that explain the apparent asynchrony. There is a ‘first-order change’ of the dot/cube, which generates two transients at every position to be kept track of by an early-level mechanism (its appearance and its departure), and there is a ‘second-order change’, which is the motion change, which does not produce any clear transient that early-level mechanisms can latch onto to create a time mark.

While there is evidence for specialised early visual sensors working at a high temporal resolution that can detect first-order temporal changes (colour, position, change in luminance), there are no such early sensors for second-order temporal changes (change in motion) (Ledgeway & Smith, 1994; Werkhoven et al., 1993). Any possible detection of second-order changes is constrained by a slower mid-level mechanism, which functions as a bottleneck on the temporal resolution of colour-motion pairings (Nishida & Johnston 2010).

By revisiting and changing the original study, Nishida and Johnston (2002) support this view, finding that second-order changes appear to be delayed relative to first-order temporal changes regardless of what features occupy these changes. A motion reversal (second-order change) would be delayed to an instant colour change (first order), but a second-order change in the direction of colour would also appear to be delayed to an instant change in position (first order). The reason temporal binding collapses at these fast rates is simply that mechanisms underpinning the detection of second-order changes are out of commission. This bottleneck persists even if we accommodate for differences in neural processing times of the relevant features (Bedell et al., 2003).

The important takeaway from the time-marker account is that this account allows for a parallel system of sensory mechanisms that fulfil their perceptual tasks in desynchronised ways while reducing or removing any effects these delays would have accurately temporally binding features according to their event time.

### 4.1.2 Dennett & Kinsbourne’s Content-sensitive Settling View

The retrodictive event time theory called the ‘content-sensitive settling view’, proposed by Dennett & Kinsbourne, (1992) builds on Dennett’s (1991) ‘multiple-drafts model of consciousness’.

Briefly put, the multiple-drafts model is a theory about how contents become cognitively accessible – i.e., available to guide reasoning, reports, deliberation, and consciously guided actions. In this view, our nervous system encodes an abundance of contents through an enormous array of sensory mechanisms operating in parallel. Yet simply processing contents does not mean that these processed contents are necessarily conscious (understood as cognitively accessible.

These contents only become part of a conscious perceptual draft when they functionally ‘grab’ the global attention of the organism’s system. Sensory contents compete to grab this attention and become part of a conscious draft (Dehaene & Naccache, 2001; Dennett, 1991).

The *content-sensitive settling view* gives a functional account couched in intentional terms to describe which contents get to grab conscious control. As its name indicates, this selection process is sensitive to contents rather than sensitive to the vehicles that represent these contents. This means that the competition value of representations is determined by how salient the contents are and how well they fit together with top-down perceptual goals, expectations, intentions, and task demands.

Any conscious draft may be retrodictively re-drafted in light of newly processed information or changing top-down goals that prompt a better perceptual interpretation of the world. On this view, a conscious perceptual experience can be re-drafted over time without us being aware of any such redrafting taking place. The re-drafted perception may completely overwrite any earlier percepts to the extent that there is no 1st or 3rd personal way to tell whether there, in fact, has ever been a first draft once it has been overwritten – the cases of retrodictive effects on temporal binding, presented above, are examples of this.

In the case of temporal binding, the contents that are temporally bound are those that add up to the representation of our environment that best lets the organism engage in some given time-sensitive behaviour or representational task (Dennett & Kinsbourne, 1992, 197). Dennett and Kinsbourne (1992, 189) take this view to be “a cheaper, less foolproof but biologically more plausible” compared to the kind of time marker view implied by Nishida & Johnston (2010). This is because, instead of our system having to keep track of explicit time markers of when each content has been processed, our system can make do with ‘sliding’ contents back and forth in time, looking for the ‘relative-phase-of-best-correspondence’ of the available contents. In this way, contents can be accurately temporally bound without requiring a system that keeps track of all temporal markers or keeps track of when features have been processed[[1]](#footnote-1). Moreover, on Dennett & Kinsbourne’s model, our brain does not need to represent the temporal properties of sensory contents twice, both when they are first encoded and then when they are consciously experienced, as they do on time marker views. On their model, our brain only represents temporal properties once when we consciously experience them—there is no need for a double transduction of temporal properties (Mölder, 2014).

So, how does one exactly specify what mechanisms underlie this content-sensitive settling? This is something Dennett and Kinsbourne (1992) are very vague about (see Akins, 1996; Mölder, 2014). I intend to correct this with the hybrid theory, which can be seen as an extension of Dennett & Kinsbourne’s model.

The explicit difference between the content-sensitive settling account and Nishida and Johnston’s time marker account is that Dennett & Kinsbourne’s account takes temporal binding to be more flexibly influenced by top-down goals and bottom-up environmental context. Temporal binding is not necessarily determined by any encoded temporal markers as in Nishida & Johnston’s view.

On one side, this flexibility allows one to account for a wider variety of retrodictive influences on temporal binding that might be difficult to explain using Nishida & Johnston’s framework. It is not clear how their theory can be extended to account for the retrodictive integration of line segment offset across hundreds of milliseconds (Drissi-Daoudi et al., 2019) or the retrodictive influence of sound cues (Stiles et al., 2018). This does not falsify Nishida & Johnston’s theory but restricts its applicability, a point they also partly concede (Nishida & Johnston, 2010, 296).

However, the explanatory breadth of Dennett and Kinsbourne’s theory is bought at the cost of explanatory depth. According to Nishida and Johnston (2010, 295), the account is too simple; they argue that “it would not distinguish, for instance, synchrony from asynchrony for repetitive changes of colour and motion regardless of the magnitude of relative delay”. In other words, it cannot explain why we perceive asynchrony—fail to temporally bind the colour and motion transients—in the colour-motion asynchrony illusion. This is because it is not clear what counts as “best” in novel lab settings and because the model provides no distinct mechanisms that one can elicit detailed predictions from[[2]](#footnote-2).

Considering the different scopes of the theories, I hold that it may be beneficial to think of Nishida & Johnston’s view as one narrow neurophysiological implementational part of Dennett & Kinsbourne’s broader functional account. In the case of keeping track of the temporal properties of 1st order transients, the brain might be employing time markers, and it might not. The important functional agreement between the two theories is that what makes up the content in the time represented is not strictly determined by the ordinality of neural processing, nor is the time represented settled millisecond-by-millisecond. The time represented is rather settled by a combination of the event time of features, the saliency of those features, and the retrospective temporal alignment of the representation of those features. It might be that Nishida and Johnston are right in positing time markers and that Dennett & Kinsbourne are right in saying that the temporal properties of the time represented are not fully determined by such time markers. But the two theories remain open about both these points. While Nishida & Johnston (2010, 296) are explicit about the fact that there are early-level mechanisms that track the onset of, for example, luminance transients, they are also openly vague about how the mid-level mechanism extracts salient features related to these time markers and how it compares these salient features across sensory modalities (See Nishida & Johnston, 2010, 296). The extraction of salient features is a selective process that is likely influenced by top-down effects, which in turn could have the consequence that the temporally bound perception is influenced by perceptual goals, tasks, intentions, expectations, etc. In the same way, Dennett & Kinsbourne (1992, 189) remain open about the idea that the perceptual system may employ time stamps (similar to Nishida & Johnston’s time markers) to temporally align sensory content “systems of "date stamps" or "postmarks" are not theoretically impossible”. The reason Dennett & Kinsbourne prefer a system without timestamps is simply that it would be less biologically taxing because it means avoiding a double representation of the temporal properties of sensory features—at the onset of processing and the level of perceptual experience.

### So, the main sentiments of both theories can be true, i.e., that the perceptual system is retrodictive and biased by top-down and bottom-up saliency clues. However, they are different in so far as they have very different explanatory scopes, which motivates the differences in the finer theoretical details. One is trying to account for the content of temporal experience in general muddy natural settings, and the other for temporal experience for the restricted case of colour-motion asynchrony.

### 4.1.3 A Neural Delay Issue for Retrodictive Event Time Theories?

When it comes to accounting for the speed-accuracy trade-off, it may seem like retrodictive theories have a problem with accounting for how our brain compensates for neural delays. This is because, according to these theories, it seems like the time of representing is delayed by the time it takes our brain to both process all the relevant signals of one event and retrospectively align these according to either time markers or their best temporal correspondence—significantly delaying our perceptual experience (see for example Moutoussis, 2012).

I think this critique is based on a misconstrual. There is no reason why our perceptual system could not rely on local perceptual processes to make the perception of simple features perceptually available as quickly as possible (just as is proposed by the standard brain time theory). While speedy perceptions of certain sensory features of an event may be inaccurate because they have not been retrodictively aligned with other features of that event, one could still report these inaccurate perceptions if probed at a time before any retrodictive redrafting has taken place. It does not follow from this view that people mistakenly take newly processed or inferred details of the same perceived event to take place at a different time (Dennett & Kinsbourne, 1992; Nishida & Johnston, 2010). Although the time of representing can come apart from the time represented according to event time theories, e.g. in the case of retrodictions, this does not always have to be the case. As such, the retrodictive event time theory does better at accounting for the compensation of discrepant processing times and does as well as the brain time theory in accounting for neural delays in the time of representing.

Despite this, retrodictive event time theories cannot account for perceptual experiences where both *the time represented* accurately reflects the relative event time of external features and *the time of representing* occurs in event time. That is, the temporally bound perceptual experience is experienced almost in event time. Take the case of gaze direction in trajectory estimation from White (White, 2018). A professional badminton player can make a shuttlecock fly at well over 350 km/h, meaning it flies at least 9 meters in 100ms. However, if the conscious percept of the shuttlecock takes more than a few hundred milliseconds to form, then the perception of the shuttlecock would lag behind the world so much that the recipient would have to return the opponent’s shuttlecock before she even saw it being hit by her opponent. However, this would create a large asynchrony between our perceptions and our actions that do not actually occur (White, 2018). The action and the perception seem subjectively to be synchronised, but this cannot be explained by the retrodictive event time theories. I argue that this kind of issue can be solved by integrating retrodictive event time theories with a predictive processing theory.

## 4.2 Predicting Time

There currently exists a range of grand unifying theoretical frameworks for explaining predictive effects on conscious perception and sensory processing more generally (Clark, 2013; de Lange et al., 2018; den Ouden et al., 2012; Dolęga & Dewhurst, 2020; Friston, 2010; Hohwy, 2012, 2013; Knill & Pouget, 2004; Nijhawan, 2008; Parr & Friston, 2017; Rao & Ballard, 1999; Summerfield & de Lange, 2014; Walsh et al., 2020; Wiese & Metzinger, 2017). At a general level of description, theories of predictive perception aim to solve the problem of how a brain is perceptually connected to its external environment when it can only ever be aware of its states and processes and how these can change and shift around. The trick of the predictive brain is to infer the nature of the signal source (the world) by detecting patterns in the varying input signal and comparing them to one’s responses. Input signals in early-level sensory processing mechanisms fuel the generation of representational models encoded at a higher level of processing in a bottom-up way. The goal of these representational models is to accurately predict the incoming signals, and as such, they can inform early-level perceptual processing in a top-down fashion. In turn, this allows the perceptual system to establish a range of generative models of the world, continuously aiming to displace incoming sensory signals with encoded predictive models.

A central reason for defending a predictive processing theory in the case of perception is that there are certain cases that predictive processing seems uniquely fit to account for—e.g., perceptual predictions of causal relations (Clark, 2013), social behaviour (Bach & Schenke, 2017; Kilner et al., 2007; Thornton & Tamir, 2020), and affordance relations (Hipólito, 2018). Such perceptions would be impossible to explain by reference to the direct processing of external stimuli due to a variety of factors, such as neural delays, desynchronised processing, impoverished stimuli, or indeterminate/ambiguity in the sensory input.

A separate reason for looking at a predictive perceptual framework is that such a system is an energy-efficient one. The better perceptual predictions a system has in place, the less energy it must spend on reorganising its internal states. Prediction errors (mismatches between incoming stimuli and predicted incoming stimuli) are costly, yet if prediction errors can be kept at a minimum, then predictions are a less costly way of implementing the kind of cognitive machinery that is required to interact efficiently with a changing external environment. At least when costs refer to metabolic expenses involved in direct processing and the updating of internal models (Clark, 2013, 195). Having encoded predictions about our external environment reduces the degree to which our perception is a blank slate to be filled through the direct processing of incoming signals. According to the radical predictive processing theorist, all that needs to ever be processed are incoming signals that clash with the predictive perception[[3]](#footnote-3).

Another important way to minimise prediction errors is by minimising errors through active inferences about how our actions would change the incoming signals (Hohwy, 2012). We can act in a certain way (including moving our bodies and consciously and unconsciously shifting our attention) to sculpt the world or at least the stimulus array to fit our predictive models and minimise prediction errors. Through ‘active inference’, the brain can predict how bodily interactions with the external world would most likely affect the input it receives, and this itself constitutes a way to minimise prediction errors across the hierarchy. Instead of just predicting environmental changes, we are also predicting how our active engagement with the environment will affect the input received and how we can act to minimise prediction errors. Dołęga & Dewhurst (2020) emphasise how attention plays a big role in this kind of prediction minimisation. Attention modulates the probability of bottom-up signals and top-down predictive inferences to determine how much influence these should have on the generation of new predictive models (Dołęga & Dewhurst, 2020). Attentional modulation can happen endogenously, coming from the system itself, by raising the probability of one or a set of predictive models to drive the system’s behaviour (e.g., searching for something specific to confirm a prediction). Or it can take place exogenously, coming from outside the system, by raising the probability of predictive models that can account for a sudden, salient, or surprising external stimulus.

We can see how predictive processing can straightforwardly be put to use to develop a predictive event time theory. As evidence from section 2.2 indicates, there are a lot of cases where it seems that predictability influences our perceptual interpretation of which features are temporally bound when. If we engage in temporal binding by exploiting predictive processing mechanisms, we can account for these cases of predictive temporal binding that are the result of predictive modelling. By having predictive models in place about how the temporal properties of different sensory signals are likely to relate, we can generate predictive perceptual hypotheses that temporally bind sensory features accordingly. Moreover, we can do this before any of these signals have finished processing.

In turn, we can develop a predictive event time theory. Such a theory would suggest that temporal binding is oriented towards predictively representing the temporal properties of external events. This means that if we have a good predictive grasp on the temporal relations of external sensory features, then we can engage in both accurate and speedy temporal binding, overcoming the speed-accuracy trade-off entirely. Engaging in temporal binding this way would thus give our perceptual system a clear advantage over a system that simply relies on direct processing, and it is also something that is likely given all the evidence presented in section 2.2 about how predictability influences temporal binding.

One big disadvantage with predictive temporally bound perceptions, however, is that such perceptions are only useful if they are accurate. It does not matter that we generate speedy temporally bound perceptions if they are misleading about how the temporal properties of external sensory features relate. If our perceptual system presents us with many such inaccuracies, this should lead us to distrust our perceptions, which would be very maladaptive. So, inaccurate predative perceptions must be kept in check. In the next section, I propose that this issue can be resolved by integrating retrodictive event time theories with this kind of predictive event time theory.

# 5 A Hybrid Model of Temporal Binding

With this rough sketch of the retrodictive and predictive framework, I will argue for two integrated hypotheses about how predictive processing can be integrated with retrodictive processing views in the domain of temporal binding. These will form the core of my hybrid theory of temporal binding.

Firstly, predictive processing allows a perceptual system to overcome the trade-off between speed of availability and accuracy of temporal binding. A perceptual system that employs predictive processing mechanisms can model the event time relations of sensory features without having to await the direct processing of every incoming signal. When integrated with the retrodictive processing theories, this solves the issue that retrodictive theories cannot account for predictive effects on temporal binding. The issue with perceptual predictions is that they are only beneficial if they are accurate, and so we cannot rely on predictions when the world changes unexpectedly. This problem, however, is solved by integrating predictive processing with retrodictive processing, as predictive errors or temporal binding inaccuracies are ironed out through retrodictive redrafting. I discuss this hypothesis in section 5.1.

Secondly, I argue that we can use the bidirectional relation that exists between the updating of predictive models and the minimisation of prediction errors to cash out the relation between the drafting and redrafting proposed by retrodictive processing accounts. This requires explicating the role attention plays in predictive processing in more detail. I hold that by understanding the role of attention within the framework of predictive processing, we can describe the influences prediction and retrodiction have on temporal binding within the same framework. This hybrid view will describe temporal binding as a highly flexible process that relies on both retrodictive- and predictive mechanisms and is determined by how our attention is directed. I discuss this in section 5.2.

There already exist two hybrid theories that emphasise the importance of both retrodictive and predictive processing, such as Grush (2005) and Hogendoorn (2022). The main sentiment for developing such a hybrid theory is also fittingly expressed by David Eagleman:

“The reason for the vigorous debate between prediction and postdiction is that both sides enjoy strong support: On the one hand, brains use every available opportunity to predict what is coming next, thus saving valuable processing time; on the other hand, experiments show that perceptions can be changed retrodictively, as may be necessitated by an unpredictably changing world.” (Eagleman, 2008, 205).

While Eagleman accurately emphasises why we need a hybrid theory, he leaves it completely open to what such a theory looks like. Grush (2005) and Hogendoorn (2022) provide two substantial hybrid theories that fit well with the first integrated hypothesis of the hybrid theory I want to propose. However, there are several differences between mine and their theories.

Firstly, Grush (2005) and Hogendoorn (2022) focus explicitly on accounting for trajectory estimation, whereas I propose a more general functional theory about how, when it comes to temporal binding in general, prediction and retrodiction are not two distinct systems working in isolation. Rather, they are two integrated processes that our brain generally exploits for the common goal of flexibly engaging in speedy and accurate temporal binding. The first integrated hypothesis I propose is, as such, intended to be broader than the trajectory estimation views of Grush (2005) and Hogendoorn (2022).

Secondly, I do not agree with Grush’s (2005) and Hogendoorn’s (2022) view that temporal experiences resemble perceptual experiences of timelines. I take it that our temporal experiences need not represent timelines sometimes, we may simply experience content that represents an isolated moment that, despite being subject to predictive drafting and retrodictive redrafting, does not constitute a timeline representation but momentous probed perceptions that depend on current attentional cues.

Thirdly, my second integrated hypothesis makes a substantially novel claim about the role attention plays in facilitating this hybrid interaction and in determining which contents we become conscious of, and which contents remain unconscious. It is from this claim that a lot of different explanations of temporal binding, predictions about the content of temporal experiences, and novel empirical claims about temporal binding follow, I set these out in sections 5.2.3 and 6.

## 5.1 Integrated hypothesis 1: Predictive drafting and Retrodictive Redrafting

In alignment with Hogendoorn's (2022) Grush’s (2005 hybrid models of trajectory estimation, I argue that the brain makes use of both retrodictive and predictive processing to engage in temporal binding at both high speeds and high levels of accuracy to effectively guide behaviour. If temporal binding was based on a system like that suggested by retrodictive event time theories (section 3.1), then we would not expect that we temporally bind sensory features in ways that compensate for the neural delays involved in sensory processing to shift the time of representing closer to event time. This is because it would be impossible to shift the time of representing to closely align with event time when that moment precedes the retrodictive temporal binding of features. That is, while retrodictive event time theories do well at accounting for how the time represented in our perceptions accurately represents the relative event time relations of external sensory features, such theories cannot account for how these perceptions could be influenced by predictive effects consciously or be available before the retrospective temporal binding has taken place. As such, they cannot account for the synchrony that exists between our perception and actions in the case of fast-paced ball games (see section 4.1.3) because, on such views, we would not perceive the relevant object long after having interacted with it. But we (as well as other animals) seem to have experienced a perception/action synchrony and thus engage in compensations for neural delays (Berry II et al., 1999; Burkitt & Hogendoorn, 2021; Kwon et al., 2015; White, 2018).

The first integrated hypothesis argues that when such retrodictive event time theories are integrated with the kind of predictive event time theory outlined above, we can solve the issue that retrodictive theories cannot account for predictive effects on temporal binding.

By integrating retrodictive theories with a predictive event time theory, we can appeal to predictive extrapolation to account for how the perceptual system compensates for neural delays. If the perceptual system displaces the direct processing of incoming stimuli for predictive models and these predictive models are accurate and action-guiding, then the temporal binding of external sensory features can take place well before direct processing and retrodictive temporal binding. Johnson et al., (2023) And Turner et al., (2024) show how extrapolation activity can both be found during early visual (retinal) processing and a series of later stages of cortical visual processing. Other evidence of predictive extrapolation comes from EEG and fMRI decoding studies of neural representations of future positions of motion stimuli (Blom et al., 2020; Ekman et al., 2017), pre-activation of neural representations of expected but absent auditory stimuli (Demarchi et al., 2019), and fMRI studies indicating how top-down expectations drive stimulus-specific activity in early visual processing (Aitken et al., 2020; Kok et al., 2016). These extrapolation mechanisms ensure that the time of representing is shifted closer to event time (Hogendoorn, 2022). Appealing to such extrapolation mechanisms allows event time theories to account for the various predictive effects on temporal binding.

For example, when added to a retrospective event time theory like the “content-sensitive settling view”, the temporal contents that are settled into conscious drafts can be said to be partially driven by predictions of the ‘relative-phase-of-best-correspondence’. The ‘relative-phase-of-best-correspondence’ would be the best predictive model of the temporal relations of external events—making them less stimulus-driven and less affected by differential neural delays.

What about the problem for the predictive event time theory sketched above? The usefulness of predictions is heavily constrained by their accuracy, and so predictive event time theories cannot account for how we engage in temporal binding when in unfamiliar temporal contexts (with a lot of novel interactions, motions, and changes). In such cases, our predictions’ accuracy is limited and needs to be kept in check by direct non-predictive processing of the unexpected stimuli. It would be dire if our system constructed time perceptual predictions that turned out to be very inaccurate because if this happened often, we would stop trusting our perceptions of temporal relations of order and simultaneity (as we cannot reliably tell which are predictive and which aren’t).

Integrating a predictive event time theory with the retrodictive redrafting of retrodictive event time theories can mitigate many of the issues that the possibility of inaccurate predictions (prediction errors) presents and help us account for the cases of retrodictive temporal binding set out in section 3.1.

If prediction errors are quickly corrected in a retrodictively revised conscious draft, then there is no big issue with prediction errors. We might even just forget them or never fully experience the erroneous predictive perceptions in any way that admits to us to consciously act on them. This is the case for perceptions of sudden motion reversals. Predictions would fail to accurately model sudden unexpected reversals, yet we perceive sudden motion reversals in close to real-time without any issues. Predictive processing allows for the real-time perception of the trajectory of an object, but this predictive trajectory perception must be overcome when the object’s trajectory unexpectedly reverses. Predictions would fail to accurately model our perception of sudden unexpected reversals as they would cause us to perceive a continued trajectory despite the reversal. It might be that this predictive model is then updated in light of incoming prediction errors, but this would involve a lag in processing the transient’s following reversal and updating our predictive model accordingly. This does not match what we perceive nor what seems to be going on in the brain; we perceive sudden motion reversals close to event time, and our brain seems to retrodictively revise previous predictions so that they seem to have predicted the reversal. Blom et al. (2021) show how, following unpredictable motion reversal, any initial latency advantages gained by predicting the trajectory are lost, and the reversal instead incurs an expected latency disadvantage. However, they also find that “the latency advantages for subsequent sequence positions were restored almost instantaneously” (Blom et al., 2021, 200). While they show that motion reversal predictions are updated to reinstate the latency advantage, this is seemingly done in a retrodictively revisionary way (e.g., by signalling motion reversals (Schwartz et al., 2007)) to ensure perceptual smoothing of the motion trajectory of the stimuli.

Even if predictive processing views could explain our experience of sudden trajectory reversals with a non-retrodictive updating mechanism that updates predictive models in light of incoming prediction errors, such a predictive event time theory could not account for the long retrodictive effects presented in section 2.1. To account for the fact that our perceptual experience of a stimulus occurring at time t1 can be retrodictively affected up to 450ms after the offset of that stimulus, we must appeal to retrodictive redrafting (Herzog et al., 2020). Take the case of perceiving an entire stream of vertical lines as being offset to some degree due to one set of lines being offset at some point in the stream (Drissi-Daoudi et al., 2019). If our system simply updated its predictions in light of prediction errors (the set of lines carrying the offset) in a non-retrodictive manner, it should only be our experience of stimuli that occur after the stimulus that prompted a prediction error that is affected by the updated predictive models. But again, this is not what happens, our perceptual experience before any stimulus that could prompt a prediction error is ‘contaminated’. The same story can be told of cases where the stimuli are rather unpredictable, yet the brain seems to compensate for differential neural processing times to represent the temporal order of sensory features in a way that better matches expectations of how external features temporally relate in event time (Alais & Carlile, 2005; Geldard & Sherrick, 1972; Hartcher-O’Brien & Alais, 2011).

By integrating retrodictive re-drafting with predictive drafting, we can explain how predictions that fail to model the world are mitigated and corrected retrodictively to ensure a continuously coherent perceptual experience of the temporally contiguous events that we need to interact with. The world does not wait until we’ve caught up, so our brain must utilise a series of tricks to represent it on the fly. By integrating retrodictive processing with predictive processing, we can naturally account for the accurate time represented yet delayed time of representing in the case of stimuli that violate prediction as such perception based on prediction errors should thus be fleeting and forgetful if there is new information or context clues available for retrodictive redrafting to take place. Erroneous perceptions of temporal relations that lead to erroneous behaviour do not have the content-sensitive competitive strength to become lasting conscious representations. In a hybrid theory, this is doubly true. Because such errors are both predictively and retrodictively ironed out.

## 5.2 Attentional Selection of Representational Content for Temporal Binding

The second integrated hypothesis of the hybrid theory concerns how we can describe the role that attention plays in determining how temporal contents get temporally bound into conscious perceptions while the temporal properties of other processed sensory contents remain unconscious. This is important to spell out because both predictions and retrodictions imply that the perceptual system is selective (i.e., it organises processed sensory information in relation to event-time expectations to guide actions). In this section, I describe how this selection takes place by appealing to the selective role attention plays in organising temporal information.

To do so, I will start by describing the inability of the standard “winning hypothesis theory of consciousness” (Hohwy, 2012) to accurately account for what it takes for content to be conscious (section 5.2.1). I then describe a theory by Dołęga & Dewhurst (2020) that integrates a predictive processing view with Dennett’s multiple draft model (Dennett, 1991) to effectively deal with these issues. This involves spelling out the role attention plays in determining how content becomes conscious (section 5.2.2). I apply Dołęga & Dewhurst’s theory to temporal binding to spell out my second integrated hypothesis and the consequences that follow from this (section 5.2.3).

### 5.2.1 Content & Consciousness

In central parts of the predictive processing literature, it is often assumed that the coalition of predictive models that become conscious is the coalition that forms the best perceptual hypothesis about the external world and determines the behaviour of the system in question (Clark, 2013; Hohwy, 2012, 2013; Otten et al., 2017; Seth, 2015). This is ‘the winning hypothesis’ theory of consciousness (Hohwy, 2012). What counts as ‘best’ depends on how probable the hypothesis is considering sensory input signals from the external world and internal signals from a hierarchy of predictive models. As has been pointed out by Dołęga & Dewhurst, (2020) this “falls short of fully accounting for conscious experiences” (for another critique, see (Marvan & Havlík, 2021) .

They argue that the winning hypothesis theory lacks a clear criterion for distinguishing between how the content of our brain’s predictive models relates to the content of our conscious and unconscious states. This is evident by how the winning hypothesis account is unable to explain cases like unconscious representation and unconscious perception. We start by exploring the case of unconscious representation.

Unconscious representations show up in binocular rivalry, where distinct images are shown to the left and right eye at the same time, resulting in the alternation between perceptual experience of one or the other. Despite only one image being consciously represented at a time, the two images are at all times equally perceptible, and so the question is why we only experience one image at a time rather than their fusion.

This is a problem for the winning hypothesis account. In the case of binocular rivalry, we have two equally probable perceptual hypotheses (e.g., a house and a face). While both inputs are processed and equally salient, only one is consciously experienced at a time (in the sense that it is accessible for reports). But if one is committed to a strong representationalism (as implied by Hohwy, 2013, 184) according to which phenomenal properties are determined and identified with representational properties, then it becomes difficult to explain why the perceptual hypotheses of both images are not phenomenally experienced at the same time. This might be explained by one of the hypotheses always marginally winning over the other, thus becoming the winning hypothesis. However, given that the content of the two hypotheses never changes, the only thing that might change is the probability assignments. So, one must account for why attribution of higher probability leads to the conscious experience of one set of contents over another. Note that simply appealing to the winning hypothesis account would be circular.

One way of accounting for this is by rejecting strong representationalism and endorsing weak representationalism so that representational properties and phenomenal properties can come apart. This opens the possibility that the non-conscious stimulus is being consciously perceived in a purely phenomenal way without it being cognitively accessible to the person enjoying the experience—they are not representational but rather purely phenomenal experiences. By distinguishing between purely phenomenal consciousness and representational access consciousness, people like Block (1995, 2007) & Lamme (2006) argue that we have phenomenally conscious experiences of much more than what we have an accessible conscious experience of. This is because the former only depends on high-capacity early sensory processing, whereas the latter depends on limited capacity higher-order processing in the prefrontal cortex—as such, we can only report a marginal aspect of our conscious experiences (Block, 2019).

Hohwy seems to adopt Block’s distinction (Hohwy, 2012, 8) and adapt it to his predictive theory:

“(i) access consciousness goes with active inference (i.e. minimising surprise through agency, which requires making model parameters and states available to control systems), and (ii) phenomenal consciousness goes with perceptual inference (i.e. minimising the bound on surprise by more passively updating model parameters and states).” (Hohwy, 2012, 7)

As such, the occasionally suppressed image in binocular rivalry would be phenomenally conscious (qualitatively experienced) by being active in some cognitively inaccessible perceptual inferential model while not being access conscious (reportable, accessible for deliberation, etc.) (see Block, 1995). This is because although the suppressed image is not accessible for reports and is, as such, not being made available to higher-order cortical systems (which, according to the quote above, is required for access consciousness), this suppressed image is still being passively updated by predictive models (which according to the quote is enough for phenomenal conscious experiences).

But it is also not entirely clear that Hohwy assents to weak representationalism as he argues that his view “seems to tick the boxes that Jackson [(2003)] and others have set out as requisites for a substantial representationalism about consciousness.” (Hohwy, 2013, 184). This indicates a commitment to strong representationalism given that one of these listed requisites is that “sensory experience plays a distinctive functional role”, a claim that proponents of phenomenal consciousness straightforwardly deny: “The controversial part is that I take P-conscious [phenomenally conscious] properties to be distinct from any cognitive, intentional, or functional property” (Block, 1995, 300).

By straddling the fence between weak and strong representationalism, Hohwy seems to vacillate between claiming that content becomes conscious just by being a highly probable perceptual inference and claiming that such perceptual inferences are insufficient for instantiating phenomenal properties (Dołęga & Dewhurst, 2020).

The latter option is untenable, as we saw above, it leads to a question of why the probability of a perception determines what we are conscious of and leaves it unanswered why we are not conscious of both perceptions in binocular rivalry cases.

The former option is untenable for multiple reasons. Firstly, saying that the updating of predictive models constitutes a phenomenally conscious state is problematic, given how phenomenal consciousness is defined. According to Block (1995, 2007) and Lamme (2006), phenomenal consciousness is defined in relation to early sensory processing that occurs in the occipital or temporal lobes, prior to any processing in later parietal or prefrontal cortices. According to Hohwy (2012) however, the updating of perceptual predictive models takes place across a large variety of processing hierarchies, and phenomenal consciousness just is a particular kind of processing rather than a specific stage of processing (Dołęga & Dewhurst, 2020). It is hard to see what Hohwy is saying is compatible with phenomenal consciousness as predictive updating takes place across the entire hierarchy of predictive models, and phenomenally non-accessible conscious states would not only include early sensory processing but also higher-level cognitive processing, but these are surely accessible and not just phenomenally conscious.

Secondly, this option is untenable because the idea of phenomenal conscious states that are inaccessible is in and of itself an idea that fits uncomfortably with empirically informed theories as it may not be empirically verifiable. This is because investigating conscious states requires eliciting some kind of report, and phenomenally conscious states do not allow for reports of any kind. Even the best empirical evidence that exists in support of phenomenal conscious experiences (according to Block (2007)) stems from experimental evidence by Sperling (1960) and Landman et al. (2003) that have been debunked by Cova et al. (2021).

For these reasons, the winning hypothesis theory of predictive processing does not provide a satisfying account of unconscious representations, and we need a satisfying account to give a clear description of what it is that determines which temporal contents we are consciously aware of when we engage in temporal binding.

Another equally worrying case for the winning hypothesis theory pointed out by Dołęga and Dewhurst (2020, 7792) is that neither of these solutions proposed for unconscious representations allows the winning hypothesis theory to account for empirically sound cases of unconscious perception. Unconscious perceptions are cases where representations can perceptually guide behaviour without any ensuing qualitative experience that can be cognitively accessed. That is, there are perceptual judgment competencies without any qualitative import. We see this in cases of blindsight (Azzopardi & Hock, 2011) where people who are phenomenologically bling (report not being able to see anything) still fare better than chance at certain visual discrimination tasks.

Since the perceptual hypotheses formed by blindsight or motion-blind patients can affect their behavioural responses just as conscious states can, these perceptual hypotheses must at some point be the winning hypothesis—as I stated, perceptual hypotheses must be a winning hypothesis to determine behaviour.

Yet the winning perceptual hypothesis that guides behaviour is never consciously experienced. As such, the winning hypothesis account is not a theory of phenomenal consciousness because to account for these behavioural effects, the winning hypothesis must say that a winning hypothesis can affect behaviour without any ensuing phenomenal states. But it is also not a complete theory of all functional aspects of access consciousness because, on this interpretation, there seems to be a winning hypothesis, in so far as there is a perception that influences behaviour just as a conscious state would, but without this hypothesis being consciously accessible. Something more is required for a representation to become conscious.

There are also non-pathological cases where a near-threshold target stimulus that is not consciously perceived is increasingly likely to be consciously perceptible if attention is postcued to the target’s location up to 100-400ms after the offset of the visual stimulus (Kentridge, 2013; Rimsky-Robert et al., 2019; Sergent et al., 2013; Thibault et al., 2016). The perceived features might even become conscious in a way that makes it seem like one was conscious of them all along.

These cases all show that we cannot differentiate between conscious and unconscious content simply by tracking sensory processing and the probability of a perceptual hypothesis. Attentional cueing also seems to be a relevant determinant. This leaves the winning hypothesis account inadequate to differentiate between conscious and unconscious states as (at least in some cases) a perceptual hypothesis can be a winning one without giving rise to a conscious state.

This problem spills over into the case of temporal binding. Unconscious perceptions of temporal properties of sensory features may influence behaviour without being temporally bound into a unified conscious percept. Consider the case of residual motion perception in ‘motion-blind’ patients (Baker et al., 1991). Azzopardi & Hock (2011) show how patients who report being phenomenally blind can still perceive the motion direction of high-contrast stimuli and that this is due to the detection of first-order motion changes in luminance rather than tracking changes in position or shapes. These people track temporal information but without binding it to any other visual cues of position or shape.

I want to claim that unless we attend to processed sensory features, we do not temporally bind their temporal properties, regardless of whether we can unconsciously act on the temporal properties of unattended processed sensory contents. The idea is that temporal binding does not occur automatically as a result of mere sensory processing. It requires attention and conscious effort to temporally bind events and sensory features together.

To flesh this idea and give an account of which processed contents get bound into conscious temporally bound perception and which do not, we need some other theoretical machinery than the winning hypothesis model. We need a theoretical framework that incorporates attention into its account.

### 5.2.2 Attention as the Difference Maker

Dołęga and Dewhurst (2020) take this approach, arguing that we can better draw the relevant distinctions between conscious/unconscious content if we combine a rigorous account of how attention influences the conscious state of representational contents (Dennett’s multiple drafts model (1991)) with the predictive processing theory. Once we have set out this account, I argue in section 5.2.3 that we can straightforwardly apply it to the case of temporal binding.

To recap, the multiple-drafts theory says that for mental states to be conscious, they must ‘grab’ influence over the perceptual system. While multiple unconscious (non-accessed) sensory contents are represented by an enormous array of sensory mechanisms operating in parallel, only a subset of these can be consciously entertained (cognitively accessed) at a time. How one of these drafts then ‘grabs’ conscious influence over the perceptual system depends on how we are probed—how our attention is directed towards processed contents.

The predictive processing framework provides us with a model of how attention can be and is deployed. It explicitly states how encoded predictions determine how attention modulates different contents or incoming stimuli, and this can be used as an implementational structure for which drafts are probed at different times, i.e., which draft we are conscious of (Dołęga & Dewhurst, 2020).

By attentional modulation, I mean that the perceptual system can deploy attentional resources to change the precision-weighting of certain contents or incoming stimuli (Hohwy, 2012, 4). Changing the precision-weighting of contents or incoming stimuli is like boosting the importance and trustworthiness of contents and stimuli, as such attention functions as an evaluation criterion for what contents should play a role in shaping our predictive perceptual models. Increasing precision by applying attention to certain inputs actively decreases the amount of prediction error of a predictive hypothesis that incorporates these inputs in the model – this, in turn, makes the predictive perceptual hypothesis more probable.

Attentional modulation can take place either exogenously or endogenously. Exogenous attention is commonly taken to be stimulus-driven, automatic and transient (Yantis, 2008). In this case, it means that exogenous attention can increase the precision of context-relevant or surprising stimuli, which is biased by ingrained saliency maps (Itti et al., 1998; Itti & Koch, 2000). Exogenous deployment of attention raises the probability of predictive hypotheses that can account for these context-relevant or surprising stimuli to become conscious. For example, a sudden noise or ingrained recognisable stimuli or pattern (such as your name or your favourite song) might bring attention to an otherwise unconscious representation.

Endogenous attention, which is commonly defined as being goal-driven, controlled, and sustained, can increase the precision of certain sensory contents that are conducive to or fit with current top-down intentions, goals, desires, and represented task demands or other kinds of goal-oriented behaviour (Aitken et al., 2020; Kok et al., 2016; Yantis, 2008). If attention is deployed endogenously, then perceptual hypotheses that explain these attentionally modulated contents become more likely to become conscious. For, if we in detection tasks endogenously attend to a Waldo-like character, our perceptual system is more likely to respond when Waldo appears in front of our eyes.

I will, as such, use the term attention to describe the mental selection of a target to guide behaviour (Wu, 2024, 306). This will, due to limits in attentional resources or limits in the resources of the mechanism relevant for attention, involve selecting certain targets at the expense of selecting others to effectively engage in specific behaviours. Endogenous attention’s selection for behaviour is set by goals, intentions, beliefs, bias, etc., while exogenous attention’s selection for behaviour is set by biased saliency maps responding to relevant salient stimuli in the environment (Wu, 2024).

By combining this predictive processing framework with the multiple drafts model’s framework, the view now states that we at all times have a bunch of contents (predictive hypotheses) at different hierarchical levels, which can become conscious through its contents being probed exogenously or endogenously. It is not enough for a draft (perceptual hypothesis) to be highly probable to become conscious, it must also be attentionally probed endogenously or exogenously. As such, the representational draft that is counted as conscious will be the one that is currently being probed by the perceptual system[[4]](#footnote-4). Given this extension of the predictive processing framework, we can now account for both unconscious representations and unconscious perceptions.

Cases of unconscious representation can be explained as cases where processed sensory content influences our behaviour but fails to become consciously accessible due to a lack of attention. In the case of binocular rivalry, we can see the rivalry as one that relies on interactions between attentional modulation and sensory processing. This is supported by studies by Zhang et al. (2011) and Li et al., (2017) who show that attention is necessary for producing rivalry. When attention is distracted, the rivalry ceases (Li et al., 2017) , and the duration in which one of the images remains dominant depends on how much it consumes attention (Zhang et al., 2013).

Cases where we unconsciously perceive something and then become aware of it (and perceive this as if we always were aware of it) can be explained as a situation where an unconsciously represented draft becomes consciously available through being postdictively endogenously or exogenously probed. Otherwise, unconscious contents get lifted into consciousness by later attentional modulation of those contents. This, for example, explains the study by Sergent et al. (2013) (see also Kentridge, 2013; Rimsky-Robert et al., 2019; Thibault et al., 2016). The reason why the perception of the presence and orientation of the visual stimulus (the Gabor patch) considerably improves is that attention retrospectively modulates what processed content we consciously access and perceive. There are also many predictive cases where attentional (endogenous or exogenous) priming of some stimulus results in an increased/decreased chance of perceiving a stimulus that would otherwise (not) have been perceived (Schacter, 1992).

Lastly, cases of unconscious perceptions that guide behaviour but remain cognitively inaccessible can be explained as contents that are inaccessible for probing by the rest of the system (Azzopardi & Hock, 2011; Baker et al., 1991)[[5]](#footnote-5). There is no cognitive access without any attentional modulation of the contents that causes them to grab control over the perceptual system. Yet this does not prevent the given contents from constituting a good predictive hypothesis at some level in the hierarchy and does not prevent it from influencing behaviour subconsciously. In cases such as blindsight, the inability to attentionally modulate these contents also explains these subjects' limited capacity for subjectively assessing the reliability of their unconscious perceptual capacities.

As such, we can better account for how content relates to conscious and unconscious perceptual experiences if we specify how attention influences which content becomes conscious.

### 5.2.3 Attention and Temporal Binding

By using the framework set out above, we can quite briefly set out the second integrated hypothesis:

Attention determines which temporal properties of some subset of processed contents become part of a conscious temporally bound perception by endogenously or exogenously modulating said contents.

This means that attention plays a selective role in temporally organising perceptions such that specific temporal properties of sensory contents are targeted to guide certain actions that depend on temporally bound perceptions. This also means that some temporal relations of sensory contents that are not selected remain consciously inaccessible to us. Our perceptual system does not integrate all sensory contents into a massively temporally bound global perception that extends across our entire perceptual field. The deployment of attention perception allows for a selective focus on a subset of targets that are conducive to guiding our behaviour—that selection for behaviour is then biased top-down by goals, intentions, tasks, expectations, etc. and biased bottom-up by saliency maps that respond to relevant, recognisable, or surprising stimuli.

I take it to be necessary for attention to interact with temporal binding in this way because while the brain processes a wealth of sensory information in parallel it is comparatively very limited in how much information it can maintain for the use of higher-order processes like thinking, reasoning, reporting, deciding, etc. (Cohen & Dennett, 2011; Dehaene, 2014; Dennett, 1991; Franconeri et al., 2007; Lau & Rosenthal, 2011; Marchetti, 2012; O’Regan & Noë, 2001). So, to guide our actions effectively, our brain must be selective about which temporal properties are being bound. According to the second integrated hypothesis, attention functions as a way to bias the perceptual organisation of temporal information, allowing only a subset of processed temporal contents to be accessible by higher-order bottlenecked processes to guide explicit cognitive and behavioural tasks. In many ways, this hypothesis reflects an assumption in the literature that attention is inextricably bound to the organisation of information in higher-order action-oriented executive processes (Baddeley, 1996; Botvinick et al., 2001; Burgoyne & Engle, 2020).

While hybrid trajectory estimation models like that of Grush (2005) and Hogendoorn (2022) give a good description of hybrid processing in trajectory estimation, it is important to spell out the selectivity in temporal binding. This is because temporal information is abundant in our environment, but to solve specific time-sensitive tasks such as playing sports or videogames, navigating traffic, hunting, and other activities that require accurately organising temporal information into synchrony and order relation, we need to be able to select which of an abundance of contents to organise to engage successfully in that task. The evidence from sections 2.1 and 2.2 clearly show how the order and synchrony relations we perceive depend on a range of bottom-up and top-down factors, and by emphasising how exogenous and endogenous attention shapes our time-perceptual predictive and retrodictive interpretation of incoming information, we can account for why these effects exist. The effects exist because attention is continuously increasing and decreasing the precision of different sensory contents in light of endogenous and exogenous bias, and in doing so, what counts as the best temporal organisation of our perceptions changes, and so do our conscious experiences. Consider some of the cases we have already discussed.

We can explain the illusion of seeing all vertical sets of lines in a perceptual stream as offset as being caused by the exogenous cueing of one set of lines that were offset (Drissi-Daoudi et al., 2019). The exogenous modulation of this offset can be explained as heightening the precision of that offset enough to enforce a retrospective redrafting of the other lines as being offset as well. A similar retrodictive explanation can be used to account for the imperceptible near-threshold Gabor patch that is more likely to be perceived when its location is exogenously cued after its offset (Sergent et al., 2013).

We can explain the cases of predictive overpowering, where the temporal order perceived is influenced by expectancies about causal patterns (Bechlivanidis et al., 2022) and audiovisual speech (Hartcher-O’Brien & Alais, 2011; Vroomen & De Gelder, 2004) , as cases where endogenous attention to causal patterns or speech predictively influences how we temporally organise the relevant inputs.

There are also cases where exogenous attention causes a retrospective redrafting of a predictive draft. In the case of motion trajectories with subsequent unexpected motion reversal, we predict to accurately perceive the relative event time of motion trajectories in event time. However, upon unexpected motion reversals that violate predictions, the exogenous cue of the motion reversal is used to signal a prediction error (Schwartz et al., 2007). This leads to an increase in the precision value of that motion reversal so that it becomes perceptually competitive with the previous trajectory prediction, thus forcing the perceptual system to retrodictively redraft our perceptions.

This second integrated hypothesis can be used to integrate a predictive event time theory with a retrodictive event time theory like the content-sensitive settling account. According to content-sensitive settling, contents are settled into a temporally bound perception at the relative phase of best correspondence, where ‘best’ is sensitive to bottom-up environmental context and top-down perceptual goals. With this predictive processing account of attentional modulation, we can understand content-sensitive settling as a consequence of how the available temporal predictive hypotheses are exogenously or endogenously probed. Moreover, when the incoming processing of sensory contents clashes with temporal bound perceptions based on a predictive hypothesis, then this content draws attention, and as attention increases the precision of this content, it forces our perceptual system to retrodictively redraft our perceptions, thus redrafting which contents are temporally bound with which.

I now turn to spell out some of the more general benefits that we gain from thinking about temporal binding in terms of these two integrated hypotheses that make up my hybrid view.

# 6 Consequences and Empirical Support

With these two integrated hypotheses in place, the hybrid theory can account for how our perceptual system makes use of selective attention to flexibly trade between engaging in predictive temporal binding to make perceptions of temporal properties available as quickly as possible and devoting precious attentional resources and time to retrodictively fine-tune the accuracy of these temporal representations. This flexibility allows the hybrid theory to explain a range of psychophysical studies on how desynchronisation in sensory processing can take place without affecting our relative event time estimation and how the brain can compensate for neural delays by shifting the time of representing closer to event time. As such, the hybrid theory is preferred to the brain time theory as it allows us to account for the psychophysical data and account for how our brain deals with the speed-accuracy trade-off.

This hybrid theory provides a substantial extension to the already existing project of Grush’s (2005) and Hogendoorn’s (2022) hybrid theories. According to the theory set out here, not only trajectory estimation but temporal binding is, in general, a hybrid task. However, the theory set forward is not merely a generalisation of these existing hybrid theories. The second integrated hypothesis adds an account of how the selection involved in retrodictive and predictive temporal binding is undertaken. This selection is undertaken by attentional modulation of sensory contents, while this may be implicitly assumed in both Grush and Hogendoorn, making this explicit is important as a series of consequences about conscious experience follow from introducing a specific selective mechanism.

I want to point out two unique aspects of the hybrid theory proposed here. The first is a consequence of how the attentional component affects our phenomenology of temporal binding; the second is a series of studies that support the hybrid theory’s claim about the central role that attention plays in temporal binding.

Concerning the first point, putting attention at the centre implies that there are attentional limits on temporal binding. This is a hypothesis that is made explicit by the hybrid theory proposed here and which needs to be put to the test. This could be done by more explicitly testing how temporal binding capacities are affected when attention is distracted or flooded. If the hybrid theory here is right that attentional modulation is necessary for temporal binding; we should do worse or even fail on simple temporal binding tasks when attentional resources are overloaded/distracted or attentional access to certain representational contents is restricted. This claim might seem uncontroversial and true, however, when it is compared to our phenomenology of temporal experiences, the claim becomes more radical.

We seem to experience everything in our perceptual field as temporally bound (across a range of otherwise dissociable mechanisms—the temporal content of our conscious perceptions seems unified and comparable across our entire perceptual field (Viera, 2020). There is no part of our conscious perceptions that has conflicting temporal content, and there are never parts of our experience that seem temporally unbound. Our temporal experience seems completely coherent at all times without any temporal gaps or temporal ambiguities about which features are temporally bound together at a specific moment. Using the words of Bayne (2010, 11), one could say we have a phenomenal field—which denotes the idea of the phenomenology of experiencing all our conscious perceptual experiences as a single unified phenomenal experience—and that this phenomenal field is temporally bound. It seems very appealing to describe temporal binding unity as a global perceptual phenomenon. This is also because it is not clear what an experience, where parts of perceptually processed features are not experienced as standing in a determinate temporal relation with all other perceptual experienced features, would feel like.

However, if the hybrid theory is right in claiming that selecting which contents are temporally bound when depends on attentional modulation, then our temporally bound perceptions should be limited by attentional capacities. However, this clashes with the phenomenological intuition that our experience presents us with a globally temporally bound perceptual field because it is impossible to fully attend to our entire perceptual field and all the processed sensory contents (Cohen et al., 2016)

To test this relation between attention and our phenomenology of temporal binding, one must elicit careful experimental setups that can elicit reports about people’s experiences of temporal binding when their attention or the temporal binding task is manipulated. If the hybrid theory is right, then people should tend to describe their experiences of temporal binding in ways that align more or less with how well they perform on temporal binding tasks – their experiences should not overflow attentional capacities.

Concerning the second point, it is important to note that while there is a lack of experimental work on the role attention plays in temporal binding, there exist a few studies that support the hybrid theory’s claim that attention plays a central role in determining the contents of temporally bound perceptions. These include how attention can increase the speed of the time of representation, improve the temporal resolution of temporal binding, and influence how likely features are to be temporally bound. While these results are compatible with Grush and Hoogendoorn’s frameworks, they are not predicted because they do not include a description of the role of attention.

Donohue et al. (2015) show how the manipulation of exogenous spatial attention, perceptual goals, and task demands have an immense effect on multi-modal temporal binding, such as the likelihood of perceiving two features as simultaneous or not. This not only indicates that temporal binding is very flexible but also that this flexibility depends on the bottom-up and top-down deployment of attention.

Relatedly Chennu et al. (2009) uses EEG to study how both the time of representing and the temporal resolution at which features can be temporally bound are largely dependent on the availability of limited attentional resources. Another study by Correa et al. (2006) investigates how paying selective endogenous attention to specific moments indicated by temporal cues positively affects temporal resolution at which we can engage in temporal binding.

McDonald et al. (2005) review how we should interpret the well-known phenomena (known as Titchener’s ‘law of prior entry’), which states that attended objects generally tend to be subjectively perceived to occur before unattended objects even when these are presented simultaneously (Titchener, 1908). Standardly, it has been thought that this effect is explained by the fact that attentional modulation increases the neural processing speed of sensory features – thus supporting the brain time theory. Yet McDonald et al. (2005) show how attentional modulation of an auditory cue influenced the perceived order of visual events without affecting event-related brain potentials, i.e., processing speeds. Instead, the perceptual shift in temporal order perception seems to be partly the result of post-perceptual (retrodictive) processing combined with the fact that attention modulates the strength of specific neural signals (see McDonald et al., 2005, 1200). Cueing attention (exogenously) with an auditory signal as such interferes with the way visually processed contents are temporally bound without affecting neural latencies. This counts against the brain time theory and in favour of the hybrid theory.

Lastly, a study by Holcombe and Cavanagh (2008) investigates the role of attention in the colour-motion asynchrony illusion. Whereas Nishida and Johnston (2002) show how the illusion is caused by the difference in the temporal structure of the feature transients, Holcombe and Cavanagh (2008) show how exogenously probing attention to these transients in the right way can make the apparent asynchrony disappear. Instead of delaying the occurrence of the colour change relative to the motion change to remove the apparent asynchrony, they only manipulate attention. Specifically, they show that when the transients caused by colour and motion are cued with a clear ring (an exogenous attentional cue), the otherwise apparent asynchrony of the transients is eliminated or reduced. Holcombe and Cavanagh (2008) argue that this is because attention gives equal rapid access to the transient of both colour and motion. This fits with Nishida and Johnston’s theory that the motion transients are neither that important nor salient as they are second-order transients that can only be detected at slow rates. The attentional cue gives access to a first-order transient of the motion at the same time at which it provides access to other superposed features (colour) that are being compared.Attention seems as such to be an important cognitive bottleneck for engaging in temporal binding tasks, as temporal binding relies on attention to temporally bind contents by accessing them simultaneously. These results cannot be predicted by the brain time theory. If sensory features are bound at some initial/early sensory processing stage, then the relative feature timing would be settled independently of how attention later accesses these features. Even if attention affects processing times, the attentional cue (the ring) should not *differentially* affect latencies of colour and motion – if affected at all, the effect should be similar for both features (Holcombe & Cavanagh, 2008). The experiment favours the explanation that the binding of features across modalities is the result of attention sampling the relevant features when cued for report – thus minimising the impact of differential latencies in early sensory processing. This is fully in line with the hybrid theory’s assumption about the influence of attention on temporal binding.

**Acknowledgements** I would like to thank two anonymous referees very much for their crucial feedback on earlier drafts of this paper, this has much improved the article and done so in substantial ways. I also want to express my gratitude to others who have given feedback or encouragement on drafts of this article or work closely related to this, including especially David Braddon-Mitchell, Alex Holcombe, and my colleagues at the Centre for Time at the University of Sydney.

**Funding** This research was funded in part by the Faculty of Arts and Social Sciences Postgraduate Research Scholarship: SC3602

**Competing interests:** The author has no competing interests to report

# References

Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R. S., Corbin, N., Callaghan, M. F., & Kok, P. (2020). Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. *PLOS Biology*, *18*(12), e3001023. https://doi.org/10.1371/journal.pbio.3001023

Akins, K. (1996). Lost the Plot? Reconstructing Dennett’s Multiple Drafts Theory of Consciousness. *Mind & Language*, *11*(1), 1–43. https://doi.org/10.1111/j.1468-0017.1996.tb00027.x

Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences*, *102*(6), 2244–2247. https://doi.org/10.1073/pnas.0407034102

Arnold, D. H., Clifford, C. W., & Wenderoth, P. (2001). Asynchronous processing in vision: Color leads motion. *Current Biology: CB*, *11*(8), 596–600. https://doi.org/10.1016/s0960-9822(01)00156-7

Arstila, V. (2015a). Defense of the brain time view. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01350

Arstila, V. (2015b). Keeping postdiction simple. *Consciousness and Cognition: An International Journal*, *38*, 205–216. https://doi.org/10.1016/j.concog.2015.10.001

Arstila, V. (2016a). *The Time of Experience and the Experience of Time*. https://doi.org/10.1007/978-3-319-22195-3\_9

Arstila, V. (2016b). Theories of apparent motion. *Phenomenology and the Cognitive Sciences*, *15*(3), 337–358. https://doi.org/10.1007/s11097-015-9418-y

Arstila, V. (2019). *Time Markers and Temporal Illusions* (pp. 339–357). https://doi.org/10.1007/978-3-030-22048-8\_18

Azzopardi, P., & Hock, H. S. (2011). Illusory motion perception in blindsight. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *108*(2), 876–881. https://doi.org/10.1073/pnas.1005974108

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312. https://doi.org/10.1111/spc3.12312

Baker, C. L., Hess, R. F., & Zihl, J. (1991). Residual motion perception in a “motion-blind” patient, assessed with limited-lifetime random dot stimuli. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *11*(2), 454–461. https://doi.org/10.1523/JNEUROSCI.11-02-00454.1991

Bechlivanidis, C., Buehner, M. J., Tecwyn, E. C., Lagnado, D. A., Hoerl, C., & McCormack, T. (2022). Human Vision Reconstructs Time to Satisfy Causal Constraints. *Psychological Science*, *33*(2), 224–235. https://doi.org/10.1177/09567976211032663

Bechlivanidis, C., & Lagnado, D. (9 C.E.). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition*, *146*, 58–66. https://doi.org/10.1016/j.cognition.2015.09.001

Bedell, H. E., Chung, S. T., Ogmen, H., & Patel, S. S. (2003). Color and motion: Which is the tortoise and which is the hare? *Vision Res*, *43*(23), 2403–2412. https://doi.org/10.1016/s0042-6989(03)00436-x

Berry II, M. J., Brivanlou, I. H., Jordan, T. A., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, *398*(6725), 334–338. https://doi.org/10.1038/18678

Berti, S. (2011). The attentional blink demonstrates automatic deviance processing in vision. *NeuroReport: For Rapid Communication of Neuroscience Research*, *22*(13), 664–667. https://doi.org/10.1097/WNR.0b013e32834a8990

Blättler, C., Ferrari, V., Didierjean, A., van Elslande, P., & Marmèche, E. (2010). Can expertise modulate representational momentum? *Visual Cognition*, *18*(9), 1253–1273. https://doi.org/10.1080/13506281003737119

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav Brain Sci*, *30*(5–6), 481–499; discussion 499-548. https://doi.org/10.1017/s0140525x07002786

Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends Cogn Sci*, *23*(12), 1003–1013. https://doi.org/10.1016/j.tics.2019.10.001

Blom, J. D., Nanuashvili, N., & Waters, F. (2021). Time Distortions: A Systematic Review of Cases Characteristic of Alice in Wonderland Syndrome. *Frontiers in Psychiatry*, *12*(603). https://doi.org/10.3389/fpsyt.2021.668633

Blom, T., Bode, S., & Hogendoorn, H. (2021). The time-course of prediction formation and revision in human visual motion processing. *Cortex*, *138*, 191–202. https://doi.org/10.1016/j.cortex.2021.02.008

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., & Hogendoorn, H. (2020). Predictions drive neural representations of visual events ahead of incoming sensory information. *Proceedings of the National Academy of Sciences*, *117*(13), 7510–7515. https://doi.org/10.1073/pnas.1917777117

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

Burkitt, A. N., & Hogendoorn, H. (2021). Predictive Visual Motion Extrapolation Emerges Spontaneously and without Supervision at Each Layer of a Hierarchical Neural Network with Spike-Timing-Dependent Plasticity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *41*(20), 4428–4438. https://doi.org/10.1523/JNEUROSCI.2017-20.2021

Changizi, M. A. (2008). The Trade-Off Between Speed and Complexity. *Behavioral and Brain Sciences*, *31*(2), 203–203. https://doi.org/10.1017/s0140525x08003853

Chennu, S., Craston, P., Wyble, B., & Bowman, H. (2009). Attention Increases the Temporal Precision of Conscious Perception: Verifying the Neural-ST2 Model. *PLOS Computational Biology*, *5*(11), e1000576. https://doi.org/10.1371/journal.pcbi.1000576

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*, *36*(3), 181–204. https://doi.org/10.1017/s0140525x12000477

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the Bandwidth of Perceptual Experience? *Trends in Cognitive Sciences*, *20*(5), 324–335. https://doi.org/10.1016/j.tics.2016.03.006

Correa, A., Sanabria, D., Spence, C., Tudela, P., & Lupiáñez, J. (2006). Selective temporal attention enhances the temporal resolution of visual perception: Evidence from a temporal order judgment task. *Brain Research*, *1070*(1), 202–205. https://doi.org/10.1016/j.brainres.2005.11.094

Cova, F., Gaillard, M., & Kammerer, F. (2021). Is the phenomenological overflow argument really supported by subjective reports? *Mind & Language*, *36*(3), 422–450. https://doi.org/10.1111/mila.12291

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. https://doi.org/10.1016/s0010-0277(00)00123-2

Demarchi, G., Sanchez, G., & Weisz, N. (2019). Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-11440-1

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548

Dennett, D. C. (1991). *Consciousness Explained*. Penguin Books.

Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, *15*(2), 183–201.

Di Luca, M., Machulla, T.-K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12). https://doi.org/10.1167/9.12.7

Dołęga, K., & E. Dewhurst, J. (2020). Fame in the predictive brain: A deflationary approach to explaining consciousness in the prediction error minimization framework. *Synthese*, *198*(8), 7781–7806. https://doi.org/10.1007/s11229-020-02548-9

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrated Neuroscience*, *9*. https://www.frontiersin.org/articles/10.3389/fnint.2015.00032

Drissi-Daoudi, L., Doerig, A., & Herzog, M. H. (2019). Feature integration within discrete time windows. *Nature Communications*, *10*(1), 4901. https://doi.org/10.1038/s41467-019-12919-7

Eagleman, D. (2008). Prediction and postdiction: Two frameworks with the goal of delay compensation. *The Behavioral and Brain Sciences*, *31*, 205–206. https://doi.org/10.1017/S0140525X08003889

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, *287*(5460), 2036–2038. https://doi.org/10.1126/science.287.5460.2036

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*, 15276. https://doi.org/10.1038/ncomms15276

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*(1), 16–25. https://doi.org/10.1016/s1364-6613(00)01568-0

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39*, e229. https://doi.org/10.1017/S0140525X15000965

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci*, *7*(7), 773–778. https://doi.org/10.1038/nn1268

Geldard, F. A., & Sherrick, C. E. (1972). The cutaneous “rabbit”: A perceptual illusion. *Science*, *178*(4057), 178–179. https://doi.org/10.1126/science.178.4057.178

Groen, O. van der, Burg, E. van der, Lunghi, C., & Alais, D. (2013). Touch Influences Visual Perception with a Tight Orientation-Tuning. *PLOS ONE*, *8*(11), e79558. https://doi.org/10.1371/journal.pone.0079558

Grush, R. (2005). Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *J Neural Eng*, *2*(3), S209-18. https://doi.org/10.1088/1741-2560/2/3/s05

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385.

Halpern, A. R., & Kelly, M. H. (1993). Memory biases in left versus right implied motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 471–484. https://doi.org/10.1037/0278-7393.19.2.471

Hartcher-O’Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. *J Exp Psychol Hum Percept Perform*, *37*(5), 1383–1395. https://doi.org/10.1037/a0024234

Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in Good Time: Long-Lasting Postdictive Effects Reveal Discrete Perception. *Trends Cogn Sci*, *24*(10), 826–837. https://doi.org/10.1016/j.tics.2020.07.001

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667. https://doi.org/10.1038/nn.4284

Hipólito, I. (2018). Perception Is Not Always and Everywhere Inferential. *Australasian Philosophical Review*, *2*(2), 184–188. https://doi.org/10.1080/24740500.2018.1552093

Hoerl, C., Lorimer, S., McCormack, T., Lagnado, D. A., Blakey, E., Tecwyn, E. C., & Buehner, M. J. (2020). Temporal Binding, Causation, and Agency: Developing a New Theoretical Framework. *Cognitive Science*, *44*(5), e12843. https://doi.org/10.1111/cogs.12843

Hogendoorn, H. (2022). Perception in real-time: Predicting the present, reconstructing the past. *Trends in Cognitive Sciences*, *26*(2), 128–141. https://doi.org/10.1016/j.tics.2021.11.003

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00096

Hohwy, J. (2013). *The predictive mind*. OUP Oxford. https://books.google.com/books?hl=en&lr=&id=3m8nAgAAQBAJ&oi=fnd&pg=PP1&dq=info:q91iSdERGM8J:scholar.google.com&ots=aDOPQOd69Z&sig=LNcd7oVGKw1KH3Cvu2EryTcLZDM

Holcombe, A. (2015). The Temporal Organisation of Perception. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization* (pp. 820–847). Oxford University Press.

Holcombe, A. O., & Cavanagh, P. (2008). Independent, synchronous access to color and motion features. *Cognition*, *107*(2), 552–580. https://doi.org/10.1016/j.cognition.2007.11.006

Hubbard, T. L. (2014). Forms of momentum across space: Representational, operational, and attentional. *Psychonomic Bulletin & Review*, *21*(6), 1371–1403. https://doi.org/10.3758/s13423-014-0624-3

Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, *44*(3), 211–221. https://doi.org/10.3758/BF03206290

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Jack, B. N., Widmann, A., O’Shea, R. P., Schröger, E., & Roeber, U. (2017). Brain activity from stimuli that are not perceived: Visual mismatch negativity during binocular rivalry suppression. *Psychophysiology*, *54*(5), 755–763. https://doi.org/10.1111/psyp.12831

Jackson, F. (2003). Mind and Illusion. *Royal Institute of Philosophy Supplement*, *53*, 251–271. https://doi.org/10.1017/s1358246100008365

Johnson, P. A., Blom, T., van Gaal, S., Feuerriegel, D., Bode, S., & Hogendoorn, H. (2023). Position representations of moving objects align with real-time position in the early visual response. *eLife*, *12*, e82424. https://doi.org/10.7554/eLife.82424

Johnston, A., & Nishida, S. (2001). Time perception: Brain time or event time? *Current Biology: CB*, *11*(11), R427-430. https://doi.org/10.1016/s0960-9822(01)00252-4

Kentridge, R. W. (2013). Visual Attention: Bringing the Unseen Past into View. *Current Biology*, *23*(2), R69–R71. https://doi.org/10.1016/j.cub.2012.11.056

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kimura, M. (2021). Prediction, Suppression of Visual Response, and Modulation of Visual Perception: Insights From Visual Evoked Potentials and Representational Momentum. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.730962

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kogai, T., Aoyama, A., Amano, K., & Takeda, T. (2011). Visual mismatch response evoked by a perceptually indistinguishable oddball. *Neuroreport*, *22*(11), 535–538. https://doi.org/10.1097/wnr.0b013e328348ab76

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*, *26*(3), 371–376. https://doi.org/10.1016/j.cub.2015.12.038

Kolers, P. A., & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Res*, *16*(4), 329–335. https://doi.org/10.1016/0042-6989(76)90192-9

Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences*, *112*(26), 8142–8147. https://doi.org/10.1073/pnas.1500361112

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends Cogn Sci*, *10*(11), 494–501. https://doi.org/10.1016/j.tics.2006.09.001

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*(2), 149–164. https://doi.org/10.1016/S0042-6989(02)00402-9

Ledgeway, T., & Smith, A. T. (1994). Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Research*, *34*(20), 2727–2740. https://doi.org/10.1016/0042-6989(94)90229-1

Lee, G. (2014). Temporal Experience and the Temporal Structure of Experience. *Philosophers’ Imprint*, *14*.

Li, H.-H., Rankin, J., Rinzel, J., Carrasco, M., & Heeger, D. J. (2017). Attention model of binocular rivalry. *Proceedings of the National Academy of Sciences*, *114*(30), E6192–E6201. https://doi.org/10.1073/pnas.1620475114

Libet, B., Wright, E. W., Feinstein, B., & Pearl, D. K. (1979). Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in man. *Brain: A Journal of Neurology*, *102*(1), 193–224. https://doi.org/10.1093/brain/102.1.193

Marvan, T., & Havlík, M. (2021). Is predictive processing a theory of perceptual consciousness? *New Ideas in Psychology*, *61*, 100837. https://doi.org/10.1016/j.newideapsych.2020.100837

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), 1197–1202. https://doi.org/10.1038/nn1512

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience*, *34*(28), 9332–9337. https://doi.org/10.1523/JNEUROSCI.1215-14.2014

Mölder, B. (2014). Constructing Time: Dennett and Grush on Temporal Representation. In V. Arstila & D. E. Lloyd (Eds.), *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality* (pp. 217–238). MIT Press.

Moutoussis, K. (2012). Asynchrony in Visual Consciousness and the Possible Involvement of Attention. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00314

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1380), 393–399.

Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *The Behavioral and Brain Sciences*, *31*(2), 179–198; discussion 198-239. https://doi.org/10.1017/S0140525X08003804

Nishida, S., & Johnston, A. (2002). Marker Correspondence, Not Processing Latency, Determines Temporal Binding of Visual Attributes. *Current Biology*, *12*(5), 359–368. https://doi.org/10.1016/S0960-9822(02)00698-X

Nishida, S., & Johnston, A. (2010). The time marker account of cross-channel temporal judgments. In B. Khurana & R. Nijhawan (Eds.), *Space and Time in Perception and Action* (pp. 278–300). Cambridge University Press. https://doi.org/10.1017/CBO9780511750540.017

O’Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. *Frontiers in Psychology*, *4*.

Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition*, *112*, 69–77. https://doi.org/10.1016/j.bandc.2016.05.002

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

Parr, T., & Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, *104*, 92–101. https://doi.org/10.1016/j.neuropsychologia.2017.08.003

Piper, M. S. (2019). Neurodynamics of time consciousness: An extensionalist explanation of apparent motion and the specious present via reentrant oscillatory multiplexing. *Conscious Cogn*, *73*, 102751. https://doi.org/10.1016/j.concog.2019.04.006

Pöppel, E. (1988). *Mindworks: Time and conscious experience*. Harcourt Brace Jovanovich.

Ramachandran, V. S., & Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature*, *304*(5926), 529–531. https://doi.org/10.1038/304529a0

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rimsky-Robert, D., Störmer, V., Sackur, J., & Sergent, C. (2019). Retrospective auditory cues can improve detection of near-threshold visual targets. *Scientific Reports*, *9*(1), 18966. https://doi.org/10.1038/s41598-019-55261-0

Ruppel, S. E., Fleming, C. N., & Hubbard, T. L. (2009). Representational momentum is not (totally) impervious to error feedback. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *63*(1), 49–58. https://doi.org/10.1037/a0013980

Rust, N. C., & Palmer, S. E. (2021). Remembering the Past to See the Future. *Annual Review of Vision Science*, *7*(1), 349–365. https://doi.org/10.1146/annurev-vision-093019-112249

Schacter, D. L. (1992). Priming and Multiple Memory Systems: Perceptual Mechanisms of Implicit Memory. *Journal of Cognitive Neuroscience*, *4*(3), 244–256. https://doi.org/10.1162/jocn.1992.4.3.244

Schwartz, G., Taylor, S., Fisher, C., Harris, R., & Berry, M. J. (2007). Synchronized firing among retinal ganglion cells signals motion reversal. *Neuron*, *55*(6), 958–969. https://doi.org/10.1016/j.neuron.2007.07.042

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology: CB*, *23*(2), 150–155. https://doi.org/10.1016/j.cub.2012.11.047

Seth, A. K. (2015). Inference to the Best Prediction. In T. Metzinger & J. M. Windt (Eds.), *Open MIND*. Open MIND. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570986

Shimojo, S. (2014). Postdiction: Its implications on visual awareness, hindsight, and sense of agency. *Front Psychol*, *5*, 196. https://doi.org/10.3389/fpsyg.2014.00196

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation. *Neuron*, *51*(5), 651–659. https://doi.org/10.1016/j.neuron.2006.08.006

Stiles, N. R. B., Li, M., Levitan, C. A., Kamitani, Y., & Shimojo, S. (2018). What you saw is what you will hear: Two new illusions with audiovisual postdictive effects. *PLOS ONE*, *13*(10), e0204217. https://doi.org/10.1371/journal.pone.0204217

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews. Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Thibault, L., van den Berg, R., Cavanagh, P., & Sergent, C. (2016). Retrospective Attention Gates Discrete Conscious Access to Past Sensory Stimuli. *PloS One*, *11*(2), e0148504. https://doi.org/10.1371/journal.pone.0148504

Thornton, I., & Hubbard, T. (2002). Representational momentum: New findings, new directions. *Visual Cognition*, *9*. https://doi.org/10.1080/13506280143000430

Thornton, M. A., & Tamir, D. I. (2020). Perceiving actions before they happen: Psychological dimensions scaffold neural action prediction. *Social Cognitive and Affective Neuroscience*, *16*(8), 807–815. https://doi.org/10.1093/scan/nsaa126

Titchener, E. B. (1908). *Lectures on the Elementary Psychology of Feeling and Attention*. Avery Publishing.

Toren, I., Aberg, K., & Paz, R. (2020). Prediction errors bidirectionally bias time perception. *Nature Neuroscience*, *23*(10). https://doi.org/10.1038/s41593-020-0698-3

Tramacere, A., & Allen, C. (2022). Temporal binding: Digging into animal minds through time perception. *Synthese*, *200*(1), 1. https://doi.org/10.1007/s11229-022-03456-w

Turner, W., Sexton, C., Johnson, P. A., Wilson, E., & Hogendoorn, H. (2024). *Progressive multi-stage extrapolation of predictable motion in human visual cortex* (p. 2024.04.22.590502). bioRxiv. https://doi.org/10.1101/2024.04.22.590502

Viera, G. (2020). The perceived unity of time. *Mind and Language*, *37*(4), 638–658.

Vroomen, J., & De Gelder, B. (2004). Perceptual Effects of Cross-Modal Stimulation: Ventriloquism and the Freezing Phenomenon. In *The handbook of multisensory processes.* (pp. 141–150). Boston Review. https://doi.org/10.7551/mitpress/3422.003.0012

Walsh, K. S., McGovern, D. P., Clark, A., & O’Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Werkhoven, P., Sperling, G., & Chubb, C. (1993). The dimensionality of texture-defined motion: A single channel theory. *Vision Research*, *33*(4), 463–485. https://doi.org/10.1016/0042-6989(93)90253-S

White, P. A. (2018). Is the perceived present a predictive model of the objective present? *Visual Cognition*, *26*(8), 624–654. https://doi.org/10.1080/13506285.2018.1530322

Wiese, W., & Metzinger, T. (2017). Vanilla Pp for Philosophers: A Primer on Predictive Processing. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. https://philarchive.org/rec/WIEVPF

Wu, W. (2024). We know what attention is! *Trends in Cognitive Sciences*, *28*(4), 304–318. https://doi.org/10.1016/j.tics.2023.11.007

Yantis, S. (2008). The Neural Basis of Selective Attention: Cortical Sources and Targets of Attentional Modulation. *Current Directions in Psychological Science*, *17*(2), 86–90. https://doi.org/10.1111/j.1467-8721.2008.00554.x

Yarrow, K., & Arnold, D. (2016). *The Timing of Experiences: How Far Can We Get with Simple Brain Time Models?* (pp. 187–201). https://doi.org/10.1007/978-3-319-22195-3\_10

Zeki, S. (2007). A Theory of Micro‐Consciousness. In *The Blackwell Companion to Consciousness* (pp. 580–588). https://doi.org/10.1002/9780470751466.ch46

Zeki, S. (2015). A massively asynchronous, parallel brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140174. https://doi.org/10.1098/rstb.2014.0174

Zeki, S., & Bartels, A. (9 C.E.). The asynchrony of consciousness. *Proceedings. Biological Sciences / The Royal Society*, *265*, 1583–1585. https://doi.org/10.1098/rspb.1998.0475

Zhang, P., Jamison, K., Engel, S., He, B., & He, S. (2011). Binocular rivalry requires visual attention. *Neuron*, *71*(2), 362–369. https://doi.org/10.1016/j.neuron.2011.05.035

Zhang, R., Kwon, O.-S., & Tadin, D. (2013). Illusory Movement of Stationary Stimuli in the Visual Periphery: Evidence for a Strong Centrifugal Prior in Motion Processing. *The Journal of Neuroscience*, *33*(10), 4415–4423. https://doi.org/10.1523/JNEUROSCI.4744-12.2013

1. In many ways, Grush’s (2005) and Hogendoorn’s (2022) trajectory estimation models of how our perceptual system keeps track of trajectories conceptually resemble Dennett & Kinsbourne’s view. All three accounts argue that our brain engages in some kind of interpretation when constructing perceptions of the temporal relations of our environment. They all agree that the time of representing and the time represented come apart. And they all propose some degree of retrospective revision of representations. [↑](#footnote-ref-1)
2. Conceptually similar models like that of Grush’s (2005)or Hogendoorn (2022) might fare better because they actually provide much more distinct predictions. [↑](#footnote-ref-2)
3. There are reasons to sceptic of such radical predictive processing accounts as there probably are quite a few hard-wired early processing mechanisms that cannot be overwritten or interfered with through top-down interventions and many top-down interventions on perceptions can seemingly be explained by a range of bias in research (see Firestone & Scholl, 2016). Moreover, it is also unlikely that the predictable parts of incoming signals are removed, and that only prediction errors ever need to be encoded, it is more likely that there are explicit neural populations for encoding best predictions and associated errors with those predictions (de Lange et al., 2018). [↑](#footnote-ref-3)
4. It is an open but perhaps empirically intractable question whether a collation of content determines the content of consciousness before the moment at which one draft of these contents is probed (see Akins, 1996). [↑](#footnote-ref-4)
5. It has also been shown in several domains how predictive error responses to visual stimuli can cause changes in predictive models without any conscious awareness of these changes: backward masking (Kogai et al., 2011), binocular rivalry (Jack et al., 2017), and the attentional blink (Berti, 2011). [↑](#footnote-ref-5)