# Binding Time

## A Theory of How Flexible and Adaptive Temporal Integration Shapes Perception

**Rasmus Pedersen**

2026

THE UNIVERSITY OF
SYDNEY

Faculty of Arts and Social Sciences
School of Humanities
Department of Philosophy
The University of Sydney

**Lead supervisor:** Prof. David Braddon-Mitchell

**Co-Supervisor:** Prof. Alex Holcombe

A thesis submitted to fulfil the requirements of the degree of Doctor of Philosophy

# Declarations

The research conducted for this thesis was supported by the award of the University of Sydney Faculty of Arts and Social Sciences Postgraduate Research Scholarship SC3602 'Future Flourishing'.

I certify that the content of this thesis is my own work. This thesis has not been submitted for any other degree or purpose. The intellectual content of this thesis is the product of my own work, and all assistance received in preparing this thesis and all sources have been acknowledged

No content produced by generative artificial intelligence tools has been used in the preparation of this thesis.

A slightly amended version of Chapter 2 of this thesis has been published as Pedersen, R. (2025) 'Against the Brain Time Theory as a General Theory of Temporal Binding', *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-025-10078-3.

A slightly amended version of Chapter 3 of this thesis has been published as Pedersen, R. (2024) 'Retrodictive and Predictive Attentional Modulation in Temporal Binding', *Synthese*. https://doi.org/10.1007/s11229-024-04787-6.

A paper based largely on Chapter 4 of this thesis is under revision at *Review of Philosophy and Psychology*.

I certify that I am the sole author of this work.

Rasmus Pedersen

'As lead supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.'

David Braddon-Mitchell

# Acknowledgements

There are many people I owe thanks to for making my three and a half years here in Australia some of the best years of my life so far.

First, thanks to my supervisor, David Braddon-Mitchell, for supporting me in innumerable ways. Thanks for all the help in making me rethink, rewrite, and review my thesis. More than anything else, thanks for providing guidance every time I encountered a new academic monster (my first tutorial, first lecture, first conference, first paper rejection, first paper revision, first job application). I appreciate how David has managed my somewhat worrisome nature. I always left our meetings with new challenges to work through and, somehow, also in a better mood.

Thanks to my supervisor, Alex Holcombe, for giving me insight into the complexity of what psychophysics has to offer and for giving me an honest and blunt impression of how messy psychological explanations are. This thesis would not have been doable without his support, as I have learned much more from him than he probably knows. I look forward to continuing our work on making science more honest and transparent.

I would also like to thank Kristie Miller for taking on the role as a third supervisor without any obligation to do so. I am grateful for the many projects she has involved me in and for inspiring me with the intensity she brings to thinking and writing about philosophy. If I can develop to have just a tenth of her productivity and tenacity, I will be more than satisfied with my efforts.

Thanks to everyone in the Centre for Time group at Sydney University for providing years of helpful comments on my sacrilegiously science-forward papers and chapters. A big thanks to the Squash Squad (Brigitte, Oscar, and Wendy) and the philosophy postgraduate cohort at Sydney University for making this so worthwhile. You will be a big reason to return to Australia.

Thanks to Luca, our beautiful cat Meatball, and our stinky cavoodle Peppie for creating such a calm and accepting atmosphere at home. I am also extremely grateful to both my parents, who have provided me with loving support through all the years of my life, even when I moved to the other side of the world.

Lastly, I owe my partner in life, Maja, more thanks than anyone for joining me on this adventure to Australia. With her genuine interest, she has forced me to rethink and clarify many of the ideas in this thesis when I have failed to properly explain them to her. But, more importantly, her seemingly infinite patience and love have made these years some I will treasure forever.

# Contents

# Abstract

This thesis develops a theory of temporal binding, the process by which we temporally integrate sensory features into perceptual experiences. Psychophysical and neurophysiological findings have revealed a remarkable amount of detail into how we engage in temporal binding. However, existing theories in psychology or philosophy do not adequately account for the adaptive and flexible nature of temporal binding, nor do they explain how our capacity to engage in temporal binding connects with our experiences of time more broadly. This thesis addresses these gaps.

In Chapter 1, I identify three explanatory problems that a theory of temporal binding must address. It must: (i) account for the many variables that influence temporal binding, (ii) address what I call the problems of neural delay and desynchronisation, and (iii) explicate the relation between the mechanisms underlying temporal binding and our apparent perceptual experiences of time.

Chapter 2 argues that so-called brain time theories, which take temporal binding to be the result of synchrony in neural processing, fail to address problems (i) and (ii). In response, I advance my own hybrid theory of temporal binding in Chapter 3. I claim that temporal integration relies on how predictive processing and retrodictive processing are modulated by attention. This theory explains (i) and (ii) by presenting temporal binding as a flexible and adaptive hybrid process that is sensitive to goals, expectations, and actions. I present a range of consequences that follow from this.

Chapter 4 addresses how the hybrid theory explains (iii) the explanatory gap between our apparent perceptual experiences of time and the dissociable and fragmented mechanisms that underlie time perception. According to the hybrid theory, our phenomenology that time is perceptually unified is imposed on the world through the attentional modulation of predictive and retrodictive processing. This explanation proposes an alternative to theories that propose the empirically implausible view that the perceived unity is inherited from unity in the mechanisms underlying time perception. While my explanation comes at the ‘cost’ of deflating the phenomenology of our perceived unity of time, these deflations are based on empirically justified arguments.

Finally, Chapter 5 presents a framework for empirically testing predictions that the hybrid theory makes about temporal binding. This involves examining three questions about the relation between how the content and our apparent phenomenology of temporal binding depend on attention. I present some hypotheses concerning these questions and new animations that can be used to develop empirical tests of these hypotheses. I end by presenting multiple ways in which we may subjectively inflate the phenomenological richness of our experiences of temporal binding.

# Prelude

This thesis investigates the phenomenon of temporal binding, the perceptual capacity by which sensory features are temporally integrated into perceptual experiences. Temporal binding enables us to perceive distinct sensory features of events as temporally related, i.e., as occurring simultaneously or in a particular order. This perceptual capacity is essential for our ability to navigate, understand, and interact with a dynamic and constantly changing environment. The phenomenon plays a central role in psychology, cognitive science, and neuroscience research on perception and conscious experience. Moreover, it is essential for answering deep philosophical questions about how temporal information of sensory features becomes perceptually available to us. Yet despite this significance, research on temporal binding, and the immense flexibility with which we integrate sensory features over time, remains remarkably sparse in philosophy. Moreover, there is no unified account of the literature from neuroscience, psychology, and cognitive science that explains how the diverse mechanisms underlying time perception give rise to our highly malleable temporal perceptual experiences. The central aim of this thesis is to integrate empirical findings from these fields to develop a functional theory of temporal binding.

To some, it may seem curious that a person with a background in philosophy is undertaking this project. Surely, this task is better left to those with formal training in one of the relevant brain sciences, as they would be ideally equipped to assess the validity of domain-specific empirical data and extract theoretical insights from them. However, providing a functional account of temporal binding is a theoretically broad task because the phenomenon is not unique to any single discipline or level of explanation. It spans multiple sciences and schools of study. It also encompasses a wide array of distinct phenomena such as the functions of different neural populations, theoretical models of the brain's sensory processing structure, psychophysical perceptual phenomena (e.g., motion blindness), different patterns and modes of attention, the perceptual role of intentions and beliefs, the validity and reliability of introspective reports of experiences, and theories of consciousness. No single person can be an expert on all these domains, and consequently, no particular discipline is exclusively responsible for studying their connections. This kind of work (unifying fragmented knowledge, resolving conceptual confusions, and identifying the underlying structure of interdisciplinary explanations) is, however, the kind of work philosophy is well equipped for because it has the conceptual tools to assess what a satisfying account would require.

This leads to the primary motivation for the project: there is a need for a unified, philosophically and empirically informed theory of temporal binding that can give a functional account of the relevant data in a way that also bridges the explanatory gap between the empirical data and perceptual experience.

A secondary motivation for this project stems from the longstanding philosophical tradition aiming to understand how conscious experience must be structured to enable the perception of temporally extended phenomena such as change, succession, and motion[1]. Philosophers interested in these questions are concerned with how we reconcile the ‘fact’ that we only ever experience the present with the ‘fact’ that we also experience temporal phenomena that do not take place only in the present, but instead over time. I argue that you cannot begin to properly address these questions about the temporal structure of consciousness without first having a somewhat firm, empirically informed understanding of how we temporally integrate sensory information. Although the thesis does not directly engage with the broader metaphysical issues surrounding the temporal structure of consciousness, it provides the theoretical groundwork for future philosophical development on these topics. By clarifying the functional mechanisms and experiential content of temporal binding, we can make more informed progress on the problem of how conscious experiences are temporally structured.

In what follows, I first expand on the primary motivation by illustrating the role temporal binding plays in our lives and providing a brief gloss of what it takes to account for this phenomenon. I then set out the secondary motivation, clarifying how my project provides a foundation for future progress on broader philosophical questions about the temporal structure of

---

[1] Philosophical investigations of the temporal structure of consciousness date back to the 4th century BC where Aristotle (Physics 4.11) discussed how our minds must be structured to perceive and experience time and temporal phenomena like change and motion. This has remained an important topic ever since, cropping up in work by Plotinus (Enneads, Volume V, III, 7), St. Augustine (Confessions bk XI, ch. XX), Boethius (The Consolation of Philosophy, BK V, Prose 6) becoming very dominant in the late 17th to 18th century with Locke (1689), Hume (1739), Kant (1787 [1929]), Thomas Reid (1785), Dugald Stewart (Stewart, 1854), Thomas Brown (1857), William Hamilton (1860), to Shadworth Hodgson (Hodgson, 1878, 1900), Robert Kelly (1882 anonymous under the alias E.R. Clay). The contemporary debate was then famously firmly grounded by William James (1886, 1890) and Edmund Husserl (1893-1917 [1991]) with James providing the basis for psychological and philosophical inquiries into the workings of the nature of temporal experience and Husserl providing the basis for phenomenological investigations into the experience of time.

consciousness. These questions will then be set aside for the remainder of the thesis. At the end of this prelude I provide a brief chapter-by-chapter overview.

### Motivation 1: Providing an Integrative Theory of Temporal Binding

We are under constant and continuous pressure to temporally organise and coordinate the activity of our body and brain relative to the flurry of external sensory information incoming from the world. To maintain a body that functions efficiently and interacts effectively with its ever-changing environment, we must be equipped with a sensory and perceptual system that enables us to track and exploit temporal relations such as temporal order and simultaneity. Whether navigating traffic, playing video games, or engaging in team sports, we rely on this temporal organisation of our perceptions. Moreover, we must do this with considerable and consistent accuracy and speed to successfully use these perceptions to guide our actions and judgments.

To achieve this, we possess a sophisticated sensory processing system that allows us to perceive an object's colour, shape, texture, taste, sound, and smell. However, merely processing and perceiving the spatial properties of sensory features is not sufficient. Basic perceptual tasks, such as determining whether two events occur simultaneously or in some specific order, require that the processed sensory features are properly temporally organised in perception. Sensory information from different modalities is processed at different times. To bind these inputs into unified objects and represent them as occurring at specific moments, we must temporally align sensory information originating from a single event so that we perceive these as occurring together. For example, to coordinate our actions relative to a ball flying through the air, we must temporally organise multiple instances of spatial processing (of the ball's colour, shape, texture, depth, luminance, contrast, and gloss) all in relation to both our own actions and the changes in other external events. Without the ability to perceptually bind the ball's features at different times and locations within its context, we would fail to perceive where the ball is at any given moment, making any effective interactions with it problematic.

This is evident when we look at cases of people suffering from the extremely rare condition 'akinetopsia' or 'motion agnosia', which results from damage to neuronal structures in the medial temporal cortex, which plays a major role in motion detection (Schenk & Zihl, 1997). While people suffering from motion agnosia are perfectly capable of spatially binding sensory features into coherent, unified objects (i.e., they can bind all the features of a ball to see a single ball), they experience changing events as a series of disjointed perceptual constellations. In severe cases called 'gross akinetopsia', dynamic events can be experienced as frozen for several seconds (Schenk &

Zihl, 1997). That is, instead of experiencing a ball flying through the air in continuous motion, people with motion agnosia see the ball spontaneously ‘rearranging’ itself at a new position along its trajectory. Not being able to temporally bind an object’s features with its motion features (i.e., seeing continuous motion) makes one worse off in even simple everyday tasks such as pouring a glass of water. It might also be downright dangerous because it decreases your ability to accurately perceive the position and movement of cars and cyclists in traffic.

Motion agnosia is an illustrative example for understanding the role of temporal binding, because it highlights several important features. Firstly, it shows how temporal binding is essential for our interaction with our environment. Secondly, it shows how temporal binding is not an all-or-nothing affair. While visual motion agnosia effectively eliminates our ability to temporally bind the visual motion features to other visual sensory features, the person’s ability to temporally bind visual features as occurring simultaneously and in temporal order is intact (though only above a certain rate of change). Thirdly, it shows how the mechanisms underlying temporal binding are dissociable. Motion agnosia is a symptom of lesions to the medial temporal cortex, but such lesions do not affect the mechanisms that allow us to temporally bind the motion to other processed features of objects through auditory cues (Devinsky & D’Esposito, 2004). In this way, it seems that our temporal perception of objects can break down in various ways, and perhaps this indicates that our otherwise perceptually unified experiences of time can come and go in degrees. In Chapters 4 and 5 I give several examples of how our temporal perceptual capacities are dissociable, fragmented, and disunified at the level of the mechanisms underlying time perception. Whereas from the perspective of our perceptual experiences, events appear to us as temporally coherent and unified as if time is perceptually represented as a single psychological kind.

Other cases beyond motion agnosia are similarly instructive for understanding the complexity of temporal binding. These come from psychophysical studies on how temporal binding is affected by all manner of interventions, contexts, and expectancies. Consider the following case called temporal ventriloquism. In a study, Hartcher-O'Brien and Alais (2011) show how two slightly asynchronous sensory signals (a sound and a flash) tend to be perceived as temporally bound (as a fused audiovisual event) in a specific, systematic manner based on the order and salience of the individual stimuli. The perceived audiovisual pair is perceptually shifted towards the more reliable salient feature (in this case, the auditory signal) regardless of whether the sound occurs before or after the flash. This means that if the auditory signal occurred before the visual signal, the audiovisual perception was perceived as occurring earlier than if the auditory signal occurred after the visual signal. As we will see in Chapters 2 and 3, there is an abundance of studies like this that supports the idea that what we perceive as temporally bound varies from situation to situation depending on what features are causally significant, of agential relevance, expected, or otherwise attentionally salient.

In Chapters 2 and 3, I explain how these empirical studies suggest that temporal binding is not completely dictated by neural processing times nor by objective timings in the external world. I suggest that temporal binding is much more flexible and that it adapts to a range of internal and external demands. Our perceptual system needs to build an accurate enough model of the world to plan and enact behaviour. However, it also cannot dawdle too much in doing so due to risks of temporally uncoupling our perceptual representations from the world. It seems our brain must flexibly deal with a trade-off between how quickly sensory signals are made consciously available and how much time it spends gathering and temporally organising information before integrating sensory signals into a conscious perception. If sensory information is made consciously available the moment it finishes processing, then signals that originate from the same event but take a different amount of time to process will not be integrated into a coherent perception. This will lead to impoverished perceptions that may fail to accurately model the world in ways needed to engage in effective planning and motor actions. However, if sensory information is always withheld until all the signals originating from one event have been processed and temporally aligned, then conscious perceptions may be so delayed that they are useless for engaging in time-sensitive behaviour. A good account must explain how our brain strikes this balance under different circumstances.

In Chapter 1, I survey neurophysiological and psychophysical findings to expand on all of the above and set out three core explanatory problems that a theory of temporal binding must address. The main project of the thesis is then to provide a functional theory that can address these problems.

### Motivation 2: Providing a Foundation to the Philosophy of Time Perception

Before I set out these explanatory problems in Chapter 1, I want to present, in some detail, my secondary motivation for undertaking this project: That theorising about temporal binding by seriously considering the empirical data is crucial for progressing with foundational philosophical discussions on the temporal structure of consciousness. Despite this, most of this philosophical literature ignores the relevant empirical problems of temporal binding.

The explanatory target of contemporary and historical philosophical theories of time perception is the question: how can we perceive temporally extended phenomena like succession, change, and motion when our conscious perception is anchored to a specific determinate moment, i.e., the present? This question has often been cashed out as a paradox between the ‘fact’ that we only ever perceive the present, and the ‘fact’ that we perceive temporal phenomena that are constituted by two temporally separated events. One option is that we perceive the two successive events themselves as present at the same time (as simultaneous). But if that is the case, then we never perceive any succession, change, or motion, we only perceive the present moment. Another option is that we perceive the events at two temporally separate moments. But in that case we never actually perceive the change, succession, motion as part of one experience, we just have two experiences that occur in succession.

The different kinds of theoretical solutions proposed to this problem can be divided into realist, anti-realist, and intermediate positions regarding how our perceptual experiences are related to the world. Realists argue that we veridically perceive the various extended phenomena either because experience itself is temporally extended, or because we have momentary experiences with contents that represent temporal extension. Anti-realists argue that we do not perceive or represent temporally extended features. That is neither our experiential processes themselves nor the representational content of our experiences directly track or mimic temporally extended phenomena in the external world. That is perception is not in the business of tracking temporal relations as they exist in the external world. Although we might believe that we have experiences as of change, motion, and succession, they argue that these beliefs are erroneous or misguided. These anti-realist and realist positions have been articulated in a range of different ways.

Most realist positions agree that our perception of temporally extended phenomena necessitates that perception has the structure of a ‘specious present’[2], a single experiential episode that perceptually appears to us as both present and as temporally extended (Dainton, 2024). Extensionalists argue that our perceptual experiences are themselves extended and that this is how we represent any temporally extended event as extended (Dainton, 2000, 2001, 2008; Foster, 1982, 1991; Hoerl, 2009, 2013; Phillips, 2008, 2011, 2014a; Piper, 2019; Rashbrook, 2013a; Rashbrook-Cooper, 2016). Retentionalists deny that experience itself is extended, taking it to be an instant or near-instant moment. Instead, they argue that we experience temporally extended phenomena because the content of our near-instant experience represents temporal extension. That is near-instant experiences simultaneously represent the immediate present, the recent past, and the

---

[2] This term was coined by William James but attributed by James (1890) to the pseudonym of Robert Kelly (1882), E.R. Clay.

expected near future (Broad, 1923; Chisholm, 1981; Gallagher & Zahavi, 2001, 2021; Grush, 2005; Husserl, 1991).

Anti-realists similarly differ in why they think that we do not have experiences of temporally extended phenomena. Philosophers like Reid (1785) argued that our experiences are like static snapshots and that these snapshots give rise to conscious experiences as of the apparent extended experiences that people erroneously claim to have[3] (for a contemporary version of a similar position see Chuard, 2017). Others who have been described as anti-realist, such as Dennett & Kinsbourne (1992), hold that the apparent experience as of temporally extended phenomena depends on our perceptual goals in a top-down fashion and depends bottom-up on the context we are in. While we may take ourselves to have experiences as of motion, change, and succession, they argue that this does not necessarily reflect anything about the temporal structure of conscious experiences. These beliefs may be the result of unreliable inferences about the temporal structure of our conscious experiences based on (fallible) introspections (Dennett & Kinsbourne, 1992). While Dennett and Kinsbourne (1992) do agree that we represent temporally extended phenomena they do not take this to mean that the contents or our experiences of such phenomena are themselves temporally extended.

Some theories lie between realism and anti-realism. These are theories such as the dynamic snapshot theory and atomism. These theories are anti-realist because they take experiences to be near-instant moments that do not represent the near-past and or near-future. They are realist because these momentary experiences give rise to experiences of succession, motion, and change (Arstila, 2016a; Le Poidevin, 2007; Lee, 2014; Prosser, 2016). While this is comparable to retentionalism, they do not appeal to memory and predictions as the way our experiences come to represent motion, change, and succession. Instead, they appeal to mechanisms that represent ‘pure’ temporal phenomenology (Le Poidevin, 2007). These mechanisms encode dynamic information like motion, change, succession, etc., and these are separate from the different mechanisms underpinning memory functions (Arstila, 2016a; Le Poidevin, 2007; Lee, 2014; Prosser, 2016). For example, these mechanisms can encode vector-like information without encoding information about changes in positions (Hassenstein & Reichardt, 1956; White, 2009, 2012). Motion can as such be neurologically encoded at an instant, which then interacts with our processing of feature locations, and together these create an experience of motion (Arstila, 2015b).

---

[3] This view is also sometimes called the cinematic model. On this kind of view, our perceptual experiences are static picture-like representations that follow each other in quick succession, much like a projector in the cinema quickly playing a series of successive pictures resulting in a smooth cinematic experience of motion, succession, and change.

Common to all these theories is a series of assumptions about how our system temporally binds sensory features and these discussions rarely include a theoretically rigorous story about the information-processing underlying temporal binding (Exceptions to this include the work by Arstila, 2016a, 2023; Grush, 2005; and Piper, 2019). I claim that any theory of how we perceive extended phenomena must build upon an empirically informed account of how we temporally bind sensory features into coherent perceptions. The idea is not that everyone interested in the perception of temporally extended phenomena should develop a unique theory of temporal binding. However, one ought to take it seriously and, as a bare minimum, avoid contradicting the psychophysical and neurophysiological data on how sensory information is processed and temporally bound. This, as we shall see, is no easy task.

Below I give four separate reasons for why adopting an explicit theory of temporal binding would be an informative and almost indispensable starting point for any philosophical theory trying to answer questions about the temporal structure of consciousness. This is intended as a means to motivate the thesis's investigation into temporal binding. I do not aim to defend any specific theory of temporal binding here. This is the object of later chapters. More importantly, I do not, in this thesis, defend any specific philosophical theory of the temporal structure of consciousness.

First, temporal binding enables us to perceive processed sensory features as parts of a coherent event located at a specific time and place, and perceiving phenomena like motion, change, or succession depends on this capacity. Understanding temporal binding requires insight into both the processing delays involved in sensory input and the mechanisms that integrate these delayed signals into the temporally bound conscious perceptions that we experience. In the following chapters, I present a range of empirical findings that challenge certain assumptions about how the perceptual system integrates temporal information across different contexts. This evidence suggests that the integration of temporal information is far from uniform and is instead highly context- and content-sensitive. As a result, any philosophical theory of our experience of temporally extended phenomena risks being empirically inadequate if it ignores these details.

Second, a theory of temporal binding informs us about what our temporal perceptual system is oriented towards representing. According to one theory of temporal binding, the brain time theory, our perceptual system is structured such that we represent sensory features that finish processing at the same time as occurring simultaneously (I discuss this view in Chapters 1 and 2). This means that if our system finishes processing a visual feature at t1 and an auditory feature at t2, we consciously experience the visual feature before the sound. The neural timing of the representational vehicles determines the temporal representational contents, and these cannot come apart. This goes against what most retentionalists think. On the other hand, you might support a theory of temporal binding according to which our perceptual system is oriented towards representing features that occur together as occurring together, despite whether these features are processed at different times (I discuss this view in Chapters 1, 2, and 3). If this kind of theory of temporal binding is right, then most kinds of extensionalism are getting something wrong, while retentionalism or some kind of anti-realism, like that of Dennett & Kinsbourne (1992) or Lee (2014), might be more accurate. This is because, according to this theory of temporal binding, the temporal properties of the representational vehicles can come apart from the temporal representational contents experienced. This contrasts with what extensionalists think. So settling what theory of temporal binding is on the right track matters for these philosophical theories of temporal perception.

Third, an empirically informed theory of temporal binding also highlights an important apparent structural difference between our perceptual experience of time and the mechanisms underlying temporal information processing. These mechanisms are very fragmented and dissociate across several dimensions of processing. Sensory information is, for example, temporally bound differently depending on the timescale, modality, and temporal judgment task in question. Despite this, our perceptual experiences present themselves to us as completely temporally unified. There is, for example, no modal distinctiveness to our temporal experiences, even though the temporal relations we perceive depend on the modalities in question.

A theory of temporal binding must account for how this experiential unity arises from fragmented and dissociable mechanisms or explain why we are misled into thinking that our temporal experiences are unified. This issue is not only central to theories of temporal binding but also to accounts of temporally extended phenomena. Consider the specious present, the hypothesised temporal window of experience within which multiple events are perceived as unified. If we perceive a colour and a tone at t1 as unified with a colour and tone at t2, are they bound by being within the same specious present? If so, how and when does the fragmented processing of features like colour and sound become unified in consciousness? Is there a late-stage mechanism that binds them at the level of experience, or do we have multiple, modality-specific specious presents that are then integrated? If the latter, how is this integration achieved? The dissociations in the mechanisms underlying temporal binding may challenge theories like extensionalism and retentionalism. Viera (2019) has, for example, argued that time perception is too fragmented to fit cleanly within any single philosophical model. Some mechanisms underlying time perception seem to align with the extensionalist principle that the temporal structure of experience mirrors the temporal structure of the represented events, while others clearly violate it. If this is right, then neither extensionalism nor views that take the extensionalist principle to be systematically false provide an accurate theory of the temporal structure of conscious experience.

Fourth, understanding temporal binding is vital for explaining how we temporally coordinate our consciously guided actions and judgments. Suppose a theory claims that the perceptual system delays fast-processing signals so they can be synchronised with slower ones before becoming conscious. This implies that conscious experience lags behind the external world by at least as much as the slowest processing delay. If true, this would constrain the degree to which conscious perception can guide timely action. It would be too slow to be fully reliable in many fast-paced contexts. Similar problems arise when considering experiences of succession and motion. Arstila (2023) argues that the specious present cannot account for the experience of temporally extended phenomena because it fails to accommodate neural processing delays. If one event is perceived at t1 (after a delay) and another at t2 (after a different delay), then the experience of succession would only arise after both delays. But that is not how we experience succession. We do not seem to first perceive two discrete events and only then experience their succession. Thus, the standard notion of a specious present seems to mischaracterise the actual dynamics of temporal perception (see Arstila 2023 for discussion and criticisms of proposed solutions). These problems are worsened by the fact that sensory processing is not only delayed but also desynchronised. Different features from the same event are processed at different speeds (see Chapter 2). This desynchronisation further challenges any model that posits a unified ‘present’ as the locus of temporal experience.

Taken together, these four considerations show why temporal binding is a foundational issue for anyone interested in the philosophy of time perception. Yet despite its importance, few philosophers have engaged directly with the problems concerning temporal binding itself. This constitutes a substantial secondary motivation for undertaking this thesis. Even for those primarily concerned with broader theoretical debates between extensionalism, retentionalism, snapshot theories, and others, understanding temporal binding is indispensable. This is partly why I dedicate this thesis solely to discussing temporal binding, setting aside the framing used in the debate between extensionalists, retentionalists, snapshot theorists, etc. The thesis will have the following structure.

## Chapter Overview

Chapter 1, *'Temporal Binding: Background and Explanatory Problems',* introduces the phenomenon of temporal binding by defining it in relation to the more well-known phenomenon of spatial binding. I also provide a range of much-needed terminology for thinking about temporal binding and set out existing theories. I then survey psychophysical and neurophysiological findings to identify three core explanatory problems that any theory of temporal binding must address.

In Chapter 2, *'Against the Brain Time Theory',* I critically discuss the brain time theory (a popular theory of temporal binding briefly outlined in Chapter 1). If this kind of theory is right, which many contemporary advocates within philosophy and the various brain sciences believe, it would provide a neat account of temporal binding. The main goal of this chapter is to argue that the brain time theory is unsatisfactory as a theory for temporal binding because it falls short of addressing two of the three explanatory problems set out in Chapter 1. This motivates looking beyond the otherwise promising brain time theory.

Chapter 3, '*A Hybrid Theory of Temporal Binding*', presents a new account of temporal binding motivated by the fact that temporal binding is affected by both retrodiction and predictions that the brain time theory cannot account for. I start by setting out two kinds of the alternative 'event time theory' (set out in Chapter 1). On the one hand, we have retrospective event time theories, which describe temporal binding as a system that retrodictively aligns sensory signals to approximate veridical perception of the timing of external events, by accommodating for the desynchronisation in neural processing times. By contrast, we can describe what I call predictive event time theories, which describe temporal binding as a system that predictively models the temporal relation of sensory features to accommodate delays in neural processing. I argue that both of these processing strategies are necessary to account for the relevant psychophysical data, but that they are insufficient on their own to address the explanatory problems set out in Chapter 1. The central aim of this chapter is to introduce and defend a novel hybrid theory. On this view, retrodictive and predictive mechanisms interact, enabling the perceptual system to dynamically make a trade-off between speed and accuracy. Moreover, I argue that attention plays a central role in not only modulating this interaction between retrodiction and prediction, but also in selecting which features are integrated into conscious temporal experience. This attentional role is crucial for explaining why temporal binding is sensitive to goals, intentions, tasks, behaviour, salient stimuli, and environmental context. I present a range of theoretical advantages of this theory, some broader theoretical philosophical consequences entailed by it, as well as evidence that uniquely supports it.

In Chapter 4, *'Deflating our Perceived Unity of Time',* I investigate the relationship between the mechanisms underlying time perception and the way in which our perception of time appears unified. The underlying mechanisms of time perception seem to be both fragmented (involving multiple distinct systems) and dissociable (can be independently manipulated). This does not fit well with our apparent experiences of time, which appear highly unified. That is, every sensory feature we experience (across modalities) seem to be part of the same temporal moment at a time and making up one unified stream of conscious experiences over time. In other words, we do not have any apparent experiences as of disunity that mirror the dissociations present in the underlying mechanisms. This presents an explanatory problem. If we take the empirical information about how our temporal information processing mechanisms dissociate at face value, then how do we account for our apparently unified temporal experiences? In this chapter, I argue that the hybrid theory provides a better explanation than other competing theories of time of how the information processing underlying temporal binding relates to our apparent perceived unity of time. A consequence of this explanation is that how our perception of time is unified is constrained by attention in distinct and important ways. This ultimately means that the different kinds of perceived unity of time should be deflated to match these proposed attentional constraints.

Finally, Chapter 5, *'How Rich is Temporal Binding?'*, sets out a new framework for how to empirically test how attention respectively constrains the phenomenology of temporal binding and our discriminatory representational capacities. It seems as if our visual experiences present all sensory features across the visual field (whether foveal or peripheral) as temporally bound in a unified present. However, empirical evidence strongly suggests that our ability to discriminate temporal relations is constrained by attentional capacities, and we cannot attend to our entire visual field at once. This creates the following tension: how can our phenomenology of the present appear globally temporally bound if our capacity for temporal binding is limited by attention? I investigate this tension by exploring two competing views concerning how the experiential richness of temporal binding may depend on attention. Either our phenomenology of temporal binding is rich across the entire visual field and overflows our limited perceptual discriminatory capacities, because only the latter are attention dependent. Or both our phenomenology and our capacities for temporal binding are attention dependent, and the reason we describe our phenomenology of temporal binding as rich and global is due to systematic subjective inflations of this phenomenology. I set out a range of questions, hypotheses, and demos of animations that can be used for future empirical studies that could settle which of these views is on the right track.

# Chapter 1

## Temporal Binding: Background and Explanatory Problems

### 1 Introduction

As described in the prelude, this thesis aims to integrate empirical findings from psychology, neuroscience, and cognitive science to develop a functional theory of temporal binding that accounts for the underlying mechanisms of temporal binding, our conscious experiences of temporal binding, and, importantly, how these are connected. This chapter sets up the background and scope of this project. I aim to present just enough for the reader to grasp the key concepts and the main explanatory targets of the thesis, leaving the bulk of the explanatory work to be done in the following chapters. In §2, I describe the phenomenon of temporal binding and the important role that temporal binding plays in temporally organising perception and in subserving a range of cognitive functions. I do this by contrasting temporal binding with the related and more widely researched phenomenon of spatial binding. In §2.1, I provide some terminology for discussing temporal binding. In §2.2, I give a brief gloss of the different kinds of theories of temporal binding which I will be critically discussing throughout the thesis. In §3, I present empirical data from psychophysical and neurophysiological studies of time perception. I do this to define three explanatory problems that a theory of temporal binding must address. These include providing a functional account of temporal binding that addresses (i) the psychophysical and neurophysiological data on how a wide variety of effects influence temporal binding, (ii) two problems I call *the problem of neural delay* and *the problem of desynchronisation*, and (iii) the relation between the mechanisms underlying temporal binding and our experiences of temporal binding. I claim that these three problems stress the need for an integrative account of temporal binding, thus setting the scope for the rest of the thesis.

## 2 What is Temporal Binding?

I use the term ‘temporal binding’ to refer to the phenomenon of how sensory features are bound into temporally integrated perceptual experiences, allowing us to perceptually discriminate whether features occur synchronously with each other or in some temporal order[4]. This notion of temporal binding can be understood as the temporal aspects of what is known as the binding problem, initially presented by the psychologists Anne Treisman and Garry Gelade (Treisman, 1996; Treisman & Gelade, 1980). Understanding temporal binding in this way also allows us to separate the temporal aspects of the binding problem from the spatial aspects of the binding problem, thus distinguishing the ‘temporal binding problem’ from the ‘spatial binding problem’.

The spatial binding problem is the problem of explaining how our conscious perception of spatial features becomes perceptually bound into separate coherent objects when our perception of these features is based on distributed and parallel neural processing. The processed spatial sensory features: colour, edges, shadows, orientation, tactile properties, etc., must be integrated and bound into spatially separate objects.

Analogously, we can understand the temporal binding problem as the problem of how the different sensory features are picked up, processed, and temporally aligned such that they make up perceptions that reflect the temporal relations in which these sensory features stand. In effect, engaging in temporal binding binds the processed sensory features into temporally distinct events.

Regardless of one’s metaphysical position on whether time and space are actual and self-subsisting (as famously defended by Samuel Clarke and Isaac Newton (In Alexander, 1956)), purely relational entities (as defended by Leibniz (In Alexander, 1956)), or subjective categories that constitute how our mind carves up the world (as defended by Kant, 1787 [1929], A23/B37–8), they are uncontroversially the most important dimensions that the brain can use to bind together multimodal sensory input and thus make sense of our complex environment.

However, it remains controversial how exactly to understand the relation between temporal binding and spatial binding. Some argue that temporal integration is a necessary prerequisite for integrating features across modalities (Engel & Singer, 2001; Van der Burg et al.,

---

[4] My use of this notion of temporal binding fits with existing uses in the psychology and neuroscience of how we perceive synchrony and temporal order (see Chen & Vroomen, 2013; Engel & Singer, 2001; Fujisaki & Nishida, 2010; Nishida & Johnston, 2010). This notion diverges slightly from another use of temporal binding as the phenomenon of subjectively compressing the temporal interval between some action and an event such that they are perceived as occurring closer together in time (See Buehner, 2010; Hoerl et al., 2020; Keane & Pylyshyn, 2006; Tanaka et al., 2019).

2008) and that in some cases, it even seems to be sufficient for such integration (Van der Burg et al., 2008, 2011; Vroomen & de Gelder, 2004). However, recent findings challenge this assumption, showing that a perception of synchrony is not necessary for certain paradigm cases of multisensory integration, such as the McGurk effect[5] (Jertberg et al., 2024). Moreover, temporal binding of sensory features seems to be influenced by a range of higher-order processing that cannot be attributed to changes in the processing of spatial features. These influences include shifts in attention (Chennu et al., 2009; Holcombe & Cavanagh, 2008; Zivony & Eimer, 2024), intentional actions (Haggard et al., 2002; Stetson et al., 2006; Stone et al., 2001), and adaptations to repetition (Di Luca et al., 2009; W. Fujisaki et al., 2004; Vroomen et al., 2004). I discuss these findings in much more detail throughout the thesis.

There exist many good accounts of spatial binding that have been pivotal in forging our understanding of how conscious experiences of spatial objects form and how these perceptions relate to the underlying sensory processing (Jackendoff, 1987; Marr, 1982; Robertson, 2003; Treisman & Gelade, 1980). Such research is important because we live in a spatial world, and so theories of spatial binding explain how we can consciously represent and interact with this world efficiently and competently. But our world is not merely spatial, it is also temporal. Therefore, having a grasp of how features of objects spatially integrate, alone, does not allow us to interact competently with the world. We also need to properly integrate the temporal properties of these features to understand how they temporally relate to each other. To get a full understanding of how conscious experiences are formed and how they relate to our underlying perceptual system, we must understand how we consciously engage with a temporal environment, which requires examining the much less studied phenomenon of temporal binding.

We have a pretty good understanding of how we engage in spatial binding. In the case of spatial binding (at least within the visual modality), the visual system receives information from retinal signals that have gone through various automatic filtering mechanisms that respond to features such as orientation, direction of motion, depth, colour, texture, etc. Over the last century, starting with Henschen (1893), incredible progress has been made in uncovering how spatial arrangement in our visual field is maintained in our visual cortex (Wandell & Winawer, 2011). One of the major achievements from this research is the ability to measure space-resolved maps in humans. This has made it possible to examine a process or signal within the human nervous system

---

[5] In the McGurk effect the presentation of a mismatched pairing of a spoken phoneme (the auditory part of a syllable) and a viseme (the visual part of a syllable), cause us to fuse the two and produce the perception of an entirely different syllable. One that doesn't match either the visual or auditory input.

and map it across different cortices, cell populations, and types of neural activity. It has been shown that the mechanisms that process these spatial features have a map-like retinotopic structure such that neurons from different cortical areas define a continuous mapping from the relative location of the visual field to the cortical surface of that area (Engel et al., 1994; Sereno et al., 1994). This retinotopic structure means that the spatial arrangement of the image hitting our retina is maintained in the primary visual cortex (V1) and a series of other extrastriate cortices (V2, V3, V4, V5), such that each spatial position present in the visual field is represented multiple times in the retinotopic maps of the various cortices (Wandell & Winawer, 2011). Oversimplifying somewhat, this means that the locations and processed spatial features encoded in these maps can be spatially bound by functionally superimposing the maps onto each other (Wandell & Winawer, 2011).

Treisman and Gelade's feature integration theory (1980) presented the first account of how the brain utilised spatial attention, synchronous neural firing patterns, and various populations of cells to undertake this spatial integration. Treisman took spatial attention to play an essential role. The encoding of objects is based on their location at an early level of sensory processing where the receptive fields are very small. Spatial attention helps to focus the cognitive processing of certain features of a stimulus at a specific location, at the cost of ignoring cognitive processing of other (distracting) features at different locations. This selective processing allows us to spatially bind the features we attend to (Treisman, 1996). In this way, spatial binding is local and limited by spatial attentional capacities. When spatial attention is diverted, overloaded, or otherwise impaired, it leads to errors in spatial binding. That is, feature maps may be incorrectly bound. For example, a blue 'X' may be illusorily bound with the colour of a spatially proximate red 'A' so that the blue 'X' is perceived as red (Treisman & Schmidt, 1982), or a series of differently coloured letters may have their features recombined, resulting in all manner of colour-letter combinations (Robertson, 2003).

Despite sometimes being framed as a solution to the binding problem, understood as a general problem (see Robertson, 2003; Treisman, 1996, 1999; Treisman & Gelade, 1980), the feature integration theory does not give us an account of how we temporally bind sensory features.

One might be tempted to think that temporal binding just relies on a temporal equivalent of the feature-mapping system used in spatial binding. On this view, just as the visual system maintains retinotopic maps that preserve the relative location of external objects and their features, the brain would maintain 'chronotopic maps' that represent when in time those objects and features occur. Temporal properties of features could then be locally encoded, compared across modalities, and ultimately superimposed with spatial feature maps to yield a unified spatiotemporal representation of the world.

However, there is no evidence indicating the existence of any such visual-temporal (or auditory-temporal, haptic-temporal, etc.) mapping system (Holcombe, 2015; Johnston & Nishida, 2001), nor is it clear what evidence of such a system would even look like. This is because it is not clear what a chronotopic map could be. Retinotopy exploits a straightforward structural correspondence between the retina and the neural representation. The two-dimensional layout of the retina provides a stable ordering of spatial relations that can be mirrored by the neural organisation of visual cortex. There is no analogous sensory array for time. We have no 'temporal retina' that simultaneously presents the temporal structure of the world in a way that could be encoded in the brain. While spatial locations are co-present on the retina and can therefore be represented in a single organised map, distinct temporal locations of incoming sensory features are not co-present. The temporal location of a moment that has passed is not available as a retinal signal, nor is the temporal location of a moment that has yet to occur. A signal that hits the retina 20 ms ago, now, and 20 ms in the future cannot all be 'on the map' simultaneously, even if they occurred in the world simultaneously. Any hypothetical spatial & chronotopic master map would therefore need to encode information in a way that is itself temporally extended, effectively being a four-dimensional structure. Yet, at any given moment, neural processing has access only to the current temporal slice of this map-like structure. The rest of the chronotopic map, the parts representing past and future, are not causally available. So when accessed, this chronotopic map only gives us access to an instant slice and is not useful for temporal integration. This problem is made multiple times more difficult by the fact that different visual features take different amounts of time to process, and by the fact that sensory information from other modalities also takes different amounts of time to process. This causes a proliferation of desynchronised chronotopic maps which need to be integrated into a unified perceptual representation. Just superimposing these maps would not help the slightest. Moreover, the temporal organisation of perception is as I will argue in Chapters 2 and 3 inherently *flexible* such that it can occur in variable ways, and also *adaptive* in that it is responsive to changing internal and external demands. What counts as the relevant temporal structure is constantly changing. Thus this kind of proposal is unlikely to succeed.

Temporal binding also differs from spatial binding insofar as it serves a distinct set of cognitive functional roles in organising cognition. White (2020) has spelled out ten different organisational functions of what he calls the perceived present (a concept functionally identical to what I refer to as temporal binding). I have compressed these into five points:

(i) Temporal binding marks an important stage of perceptual processing in which sensory information becomes consciously and perceptually available and remains accessible for some amount of time (White, 2020).

(ii) Information in temporally bound perceptions may be used by other processes to support cognitive deliberation, consciously guided temporal coordination of actions, and learning, and may also be transferred to long-term memory or maintained for a longer period in working memory (White, 2020).

(iii) Temporal binding supports a large capacity for the global perceptual integration of features, objects, events, and contextual information (White, 2020). Some claim (Zimmermann et al., 2014) that the amount of information, complexity, integration, and number of items integrated is very high in this perceptual moment. Others, myself included, are sceptical about the capacity of information integrated and stored at this moment due to limitations of attention and working memory (Cohen et al., 2016; Franconeri et al., 2007; Jacob et al., 2013; Öğmen et al., 2013; Pylyshyn & Storm, 1988; Yantis, 1992). I discuss this in detail in Chapters 4 and 5.

(iv) Temporally bound perceptions include semantic information about emotional expressions (Kiss & Eimer, 2008), motion information about otherwise unobservable vectors (Hassenstein & Reichardt, 1956), and kinetic variables such as force, mass, causality (White, 2009, 2012), and trajectories (Grush, 2005; Hogendoorn, 2020).

(v) It is often assumed that temporal binding is a continuously operating conscious process[6] that allows for a coherent and dynamically stable perceptual representation of the world across multiple timescales (Akyürek, 2025; Hogendoorn, 2022; Mégevand et al., 2013).

I now provide a more precise vocabulary to help differentiate theories of temporal binding.

---

[6] Note, however, that whether temporal binding is always a conscious phenomenon is still an open question. It is possible that we can engage in temporal binding of sensory information outside conscious awareness. This remains underdetermined by available data, and settling on an answer prematurely risks begging the question against some theories of temporal binding and some theories of consciousness. This is due to a long and rich, controversial debate about what it means for visual perceptions (as well as other types of perception) to count as conscious rather than unconscious (Seth & Bayne, 2022). In this chapter, I remain silent on exactly what functional role consciousness plays in our ability to engage in temporal binding. However, this issue will be part of the ongoing discussion in Chapters 2, 3, 4, and 5.

### 2.1 Terminology for Different Kinds of Time

Theoretical discussions of time perception can be quite confusing due to the ambiguity of the many possible referents of the word ‘time’. To avoid misunderstandings, we need a proper vocabulary for talking about these different referents. We should at least distinguish between the important fine-grained distinctions of the three broader terms: subjective time, event time, and brain time.

*Subjective time* is a vague term that refers to our perceptual experiences of time, e.g., ‘I saw the flash of lightning way before I heard the thunder.’ More specifically, subjective time can refer to either the *time of representing,* which is the time of occurrence of the perceptual experience, or the *time represented,* which is the temporal representational content of the perceptual experience. It is obvious to many that the act of representing and the representational content come apart in the spatial case, e.g., the representation of a blue chair is not itself blue or chair shaped. However, the answer is less obvious in the temporal case. A host of people take the ‘time of representing’ and the ‘time represented’ to necessarily refer to the same time, arguing that they cannot dissociate. They often claim that this is because the temporal properties we attach to represented sensory features (the time represented) are inherited from or mirror the time at which that perceptual experience of the features is realised (the time of representing). Some also argue that there are strong neurophysiological reasons to believe that the time represented and the time of representing do not come apart (Arnold et al., 2001; Arstila, 2016a; Moutoussis, 2012; Zeki, 2015; Zeki & Bartels, 1998). Others disagree, arguing that there is nothing special about the temporal case and that these can dissociate freely (Dennett, 1991; Dennett & Kinsbourne, 1992; Grush, 2005; Holcombe, 2015; Johnston & Nishida, 2001; Lee, 2014; Nishida & Johnston, 2010). I discuss different interpretations of how these may relate in Chapters 2, 3, and 4.

*Event time* refers to the time at which events take place in the external world, e.g., the thunder occurred at the same event time as the lightning. Event time can be considered an objective measurable moment in time, at least given the frames of reference used in this context.

*Brain time* refers to the time at which neural processes take place in the brain, e.g., the flash of lightning was processed in the early visual cortex V1 at t1, while the sound of thunder was processed in the early auditory cortex A1 at t2. More specifically, within the context of the discussion of temporal binding, brain time is a vague catch-all term that can refer more precisely to the *onset of sensory signals*, *content fixation,* or *the perceptual endpoint*. The onset of sensory signals is the moment when a stimulus first impinges on a neural mechanism, e.g., light travels faster than sound, so the onset of the flash of lightning will be earlier in event time than the onset of the thunder. Content fixation is the moment a neural mechanism finishes its processing (has transduced the relevant input) such that it is available to take part in a conscious perceptual representation (though does not necessarily do so), e.g., a flash could finish processing in V1 at t1 yet be consciously perceived at t2. The perceptual endpoint is the moment when a neural mechanism finishes its processing of some sensory feature and reaches the process responsible for conscious perceptual experience, e.g., if the flash and the sound both reached their respective perceptual endpoints at t1 they would be experienced as occurring at t1. Both the nature and definition of perceptual endpoints are, as I discuss in Chapters 2 and 3, very debatable. This is because their existence and definition depend on one's underlying theory of sensory information processing, one's theory of consciousness, and the assumption that there is always a determinate point in time at which a sensory feature becomes conscious (an assumption I will contest).

Most of the theoretical discussions of what temporal binding amounts to and how temporal binding is experienced can be understood as disagreements about how these notions should be cashed out, how they relate to each other, and which mechanisms underpin them.

## 2.2 Theories of Temporal Binding

Existing theories of temporal binding can be roughly divided into two kinds: brain time theories and event time theories. These two theoretical frameworks differ in terms of what they take to determine temporal binding. While these two kinds of theories come in many varieties, this distinction shapes the primary divide in the literature.

### *2.2.1 Brain time theories*

Brain time theories take temporal binding to be a perceptual process that is fully determined by perceptual endpoints. If two sensory signals, say a sound and a flash, reach their individual perceptual endpoints at the same time, then these two features are experienced as occurring simultaneously in subjective time, regardless of when these sensory features occur in the world. It follows then that the time represented stands in an isomorphic relation to the time of representing, because both of these are determined by perceptual endpoints. As Arstila (2016a, 168 (my brackets)) puts it: '[...] the order in which stimuli are experienced to occur is the same in which the processing related to contents is completed' (Arstila, 2016a, 168). This means that the distinction between 'the time represented', 'the time of representation' and 'perceptual endpoints' is merely conceptual. By fixing a definition of perceptual endpoints, one automatically fixes the time of representing and the time represented (I present this in more detail in Chapter 2).

This definition of the brain time view leaves a lot of room for different brain time theories due to the many possible interpretations of perceptual endpoints. Some claim that processing is a very modular and feedforward process (Arnold, 2010; Zeki, 2007; Zeki & Bartels, 1998), while others claim that sensory processing is also importantly non-linear, and so determining when a feature is processed depends on a range of interactions between feedforward and recurrent processing (Arstila, 2015, 2016a, 2016b; Lamme, 2006; Lamme & Roelfsema, 2000; Piper, 2019; Scholte et al., 2008). I critically discuss these views in detail in Chapter 2[7].

### *2.2.2 Event time theories*

By contrast, event time theories take temporal binding to be largely oriented towards temporally binding sensory features into coherent perceptions relative to when they occur in the external world (relative to sensory features' event time). Our perceptual system is as such oriented toward accurately representing the event time of external features, with delay mechanisms and other compensation mechanisms in place to approximate veridical event time perception. If two sensory features occur at the same event time, then we should also (if our system is functioning optimally) perceive them as occurring synchronously in subjective time (time represented), regardless of whether the content fixation of the sound stimulus occurs before/after the content fixation of the flash stimulus. As such, the time represented can come apart from the time of representing.

Event time theories also differ in spelling out how temporal binding is achieved. On one view, the perceptual system has a special affinity for simultaneity constancy and utilises complex memory mechanisms to align signals and retain event time relations in perception (Kopinska & Harris, 2004). Libet (1978) is famous for advocating this kind of view. He argued that (a) there are neuronal processes responsible for marking the onset of stimuli, and (b) there is an automatic process that refers back to these time markers when, after the necessary processing delays, a subjective conscious experience is formed. Nishida & Johnston (2002, 2010) present a distinct version of this view, according to which we have intra-modal early sensory mechanisms with high

---

[7] This characterisation of the brain time theory also leaves room for different understandings of how the perceptual system is implemented in our nervous system. To some, temporal binding is the literal synchronisation of oscillations in neural firing patterns (Crick & Koch, 1990; Engel & Singer, 2001). Others argue that there is a complex of local synchronic and diachronic resonance patterns that modulate each other to temporally coordinate sensory processing (Prinz, 2012), or that oscillatory patterns structure the specific length of integration windows such that features processed within the same integration window are perceived as occurring synchronously (Piper, 2019). I do not cover these theories of neurophysiological implementation in the thesis. This is because, as I argue in Chapter 2, there are good reasons to be sceptical about the brain time theory's capacity to provide a satisfying theory of temporal binding at a functional level.

temporal resolution that mark the time of sensory feature onset, and inter-modal mechanisms that use these time markers to temporally align sensory features across modalities.

Dennett & Kinsbourne (1992) present an event time theory without time markers. They argue that accurate event time estimation could be ensured if our perceptual system is just sensitive to the content of processed sensory information and the context at hand. Instead of keeping track of the onset of sensory signals, our perceptual system simply 'slides' represented contents back and forth relative to each other, opting for the 'relative-phase-of-best-correspondence', i.e., the best temporal fit between these contents. They explain temporal binding as functioning in similar ways to a filmmaker who is tasked to synchronise a soundtrack relative to a video track without relying on any time markers (Dennett & Kinsbourne, 1992).

I discuss these views much more in Chapter 3, where I will provide a new theory of temporal binding that draws inspiration from these event time theories.

## 3 Three Explanatory Problems From Psychophysical and Neurophysiological Findings

I now turn to provide a brief overview of the relevant empirical literature from psychophysical and neurophysiological studies on time perception. I do this to draw out what I take to be the three most important explanatory problems that a theory of temporal binding must address. These set the scope and explanatory project for the rest of the thesis.

### 3.1 Problem (i): The Many Effects on Temporal Binding

The first explanatory problem concerns accounting for all the data from psychophysical studies on how temporal binding can be influenced. This is a difficult thing to account for, both because of the variety of interventions, contexts, stimulus combinations, and timescales that influence temporal binding, and because of the internal difficulties in interpreting what the data reveal about the nature of temporal perceptual experience itself.

It is important, when analysing psychophysical data, to distinguish whether a study investigates effects on *perceptual capacities* or *judgment capacities* (Firestone & Scholl, 2016). That is, one should keep track of whether a study is testing how a participant directly senses some state of affairs, versus testing how a participant infers something about some state of affairs. If a person sees the colour of a white shoe (only by looking at it) and uses this information to report that it is white, then this is a clear case of perception. In comparison, a person who looks at a shoe and uses this information to report that it 'looks' expensive is a clear case of judgment. The latter is a judgment because expensiveness can only be assessed based on a range of inferences that go beyond the sensory information gleaned from looking directly at a shoe (unless it has a price tag attached). Whereas the colour of the shoe can be both directly perceived and also subject to judgment (Firestone & Scholl, 2016). For example, if you know that you are in a room filled with red light and a white shoe, you may judge the shoe as white even though it is perceived as red. Firestone and Scholl (2016) argue that the distinction between perception and judgment is clear and intuitive, and that we should opt for explaining most of what appears as top-down effects on perception as top-down effects on our judgments (Firestone & Scholl, 2016, 9-10). However, as pointed out by Keller (2016) the intuitive clarity of this distinction falls apart once applied outside the visual modality:

> ‘When we broaden the meaning of perceiving to include all sensory modalities, it is much less obvious whether the uncomfortableness of the shoe is judged or perceived. If wearing the shoe hurts, do we perceive that it is an uncomfortable shoe, or should that also be considered a judgment because it involves an extra step from perceiving the pain to judging the shoe to be uncomfortable? If the left shoe hurts more than the right shoe, do we perceive the difference, or is anything that involves a comparison a judgment? The answers to such questions are not clear and intuitive.’ Keller (2016, 38).

It is even the case that the distinction between perception and judgment can be unclear intra-modally. Keller (2016) points to a series of studies showing that emotional valence is the most important factor when it comes to olfactory perceptions (see Haddad et al., 2008; Soudry et al., 2011; Stevenson, 2010). According to such studies, olfaction is largely an evaluative sense, thus making olfaction a paradigm case where perception and judgment are inseparable.

The reason for challenging this distinction in these cases is not to argue that there is no such thing as the distinction. There might in fact be ways to tease them apart. For example, there is clearly a conceptual difference between the following two information processing chains: (a) sensory information of a pain stimulus is processed. This causes a belief about the sensory information (judgment), which is then decoded as an experience of pain (perception); or (b) sensory information of a pain stimulus is processed. This causes an experience of pain (perception), which then generates a belief about the sensory information (judgment).

The fact that we can make such conceptual distinctions, however, does not mean that they could always be plausibly untangled empirically or functionally. In some cases, the distinction between perception and judgment may simply be a distinction without a difference.

I think this is at least so in the case of temporal binding. I will argue that it is impossible to draw a clear functional or empirical distinction between our reports and experiences of temporal binding as instances of perceptions or judgments. When multiple sensory features, even within one modality, are temporally bound together into one ‘event’, it is far from clear whether top-down effects that influence us to report these features as occurring simultaneously are due to an effect on perception or judgment. This is partly because temporal binding can, within very short timescales, be influenced by both top-down effects (goals, intentions, expectations) and bottom-up effects (salience, causal regularities, and exemptions thereof). This should make us wary about whether any distinction between perception and judgment is plausible and/or useful in the case of temporal binding. This matters because, when we talk about effects influencing our experience of temporal binding, it is hard if not practically impossible to disambiguate whether such an experience is solely the result of perception or also partly the result of an inferential judgment.

In a series of psychophysical studies by Haggard et al. (2002) Cunningham et al. (2001), and Stetson et al. (2006), we see clearly how top-down inferences, such as intending to cause a specific event to occur, can affect your perception/judgment of the temporal relation between your actions and some succeeding or simultaneous ‘effect’ of that ‘cause’. In the paradigm case, participants are presented with a clock face and asked to judge the onset time of different events. In one condition, the participant was asked to voluntarily press a key at their convenience, which was then followed by a tone 250 ms later. In the condition where the participant voluntarily pressed the button, the action was perceived as occurring later (closer to the tone), and the tone was perceived earlier (closer to the action) compared to an involuntary condition. Are participants reporting the result of a perception that is influenced by their involved agency, or are they making a judgment about the timing of the two events based on the inference that they caused the event? No interventions or reports allow for any clear interpretation of this.

As I will explore in depth in the ensuing chapters, many other cases of temporal binding are the result of both top-down and bottom-up effects. These include causal expectations (Bechlivanidis et al., 2022; Fornaciai & Di Luca, 2020), adaptations to repetition (Di Luca et al., 2009; Waka Fujisaki et al., 2004), surprising violations of adaptations (Stetson et al., 2006), one’s temporal judgment task (Love et al., 2013), representational momentum (Hubbard, 2014), audio-visual speech recognition (McGurk & MacDonald, 1976), modality pairings (Chen et al., 2013; Hartcher-O’Brien & Alais, 2011), arm-crossing (Yamamoto & Kitazawa, 2001), and many retrodictive effects (stimuli occurring later affect how we temporally bind features occurring earlier) (Drissi-Daoudi et al., 2021; Herzog et al., 2020; Sergent et al., 2013; Shinsuke Shimojo, 2014).

Given the variety of ways temporal binding may be influenced at a variety of timescales, both retrospectively and predictively, ranging from ten to almost five hundred milliseconds (Di Luca et al., 2009; Herzog et al., 2020), it is likely impossible to make a clear distinction between which cases of temporal binding are a consequence of perception and which are a consequence of judgment. These considerations, along with the different effects on temporal binding noted above, lead us to the first explanatory problem:

> ***Problem (i):*** A theory of temporal binding must account for the psychophysical empirical data on how a wide variety of effects can influence our experience of temporal binding.

However, such a theory need not account for these effects on experience by clearly distinguishing whether these are effects on perceptions or effects on judgments. I make it clearer in Chapters 2 and 3 why this is so, by arguing that temporal binding depends both on direct perceptual processing, predictive processing, retrodictive inferences, and how our attention interacts with one's relevant context, behaviour, intentions, etc.

### 3.2 Problem (ii): The Problem of Neural Delay and the Problem of Desynchronisation

The second explanatory problem I want to emphasise concerns what I call the problem of neural delay and the problem of desynchronisation.

The problem of neural delay is the problem of explaining the fact that sensory processing takes time, yet our perception of multi-modal events appears to be immediate and sometimes even aligned with real-time events (Eagleman, 2008; Hogendoorn, 2022; Johnson et al., 2023; Nijhawan, 2008). If taken at face value, this seems impossible due to the time it takes for external sensory signals to propagate and reach our peripheral nervous system, and the various neural delays involved in sensory processing and transmission of such information through the relevant cortical areas. Even if neural processing is incredibly quick, it will be the case that the external world has changed by the time some sensory signal has been processed and made consciously available. Moreover, sensory processing can be slow and takes place across multiple stages. A question that arises from this is: if our perceptual experience of objects and events lags behind event time by such a substantial amount, then why do we sometimes not notice this delay, and why are such delays not detrimental to the coordination of a range of sensorimotor tasks that are guided by conscious perception? There is, as such, an explanatory problem of explaining how our perceptual system deals with delays in neural processing time in order to perceptually guide actions and judgments, and to present us with perceptions that make it appear that we represent the world in real time. I will call this the problem of neural delay (inspired by Hogendoorn, 2022). I present this problem in more detail in Chapter 2.

The second problem, the problem of desynchronisation, refers to the problem of explaining how we temporally bind sensory features into coherent perceptions such that we perceive the synchrony and temporal order relations of sensory features, even when the processing of these features takes different amounts of time. While our perceptions are integrated across modalities, the processing of the various sensory features that make up these perceptions desynchronise at several stages. The features originating from one event desynchronise as they propagate, hit our peripheral nervous system, are processed in parallel, and as they pass up and down the processing hierarchy (Bullier, 2001; Groen et al., 2013; Holcombe, 2009; Pöppel, 1988; Zeki, 2015; White 2023). I present the details of this desynchronisation in more detail in Chapter 2. There is, as such, an explanatory problem of explaining how our perceptual system manages to temporally bind features that occur synchronously in the external world, when the processing of those features is asynchronous. I will call this the problem of desynchronisation (Hogendoorn, 2022).

These two problems hint at the trade-off I mentioned in the prelude, between how quickly sensory signals are made consciously available and how much time is spent gathering information before integrating sensory signals into one coherent, conscious perceptual moment. We need to take in information to form an accurate perception of the world that allows us to perceive which features occurred when. However, sensory information also needs to be rapidly perceptually available to avoid too significant delays relative to when the actual event occurred. This gives us the second explanatory problem.

> ***Problem (ii)***: A theory of temporal binding must provide an empirically and phenomenologically accurate solution to the problem of neural delay and the problem of desynchronisation. This involves answering how our perceptual system temporally organises sensory information in the face of neural delays and desynchronisation.

In Chapter 2, I argue that the brain time theory cannot address this explanatory problem. I use this to motivate the need for a different theory of temporal binding altogether, a theory I present in Chapter 3.

### 3.3 Problem (iii): Relating the Mechanisms to Experiences of Temporal Binding

The third and final explanatory problem concerns how the information processing mechanisms posited to underlie temporal binding (and temporal perception more generally) relate to the phenomenology of our temporal experiences. This is particularly challenging because, while our experiences of time appear coherent and unified, evidence from neurophysiological, pharmacological, and psychophysical studies indicates that the mechanisms underlying temporal perception are highly fragmented and dissociable. It is therefore important to explain how our perceptual experiences connect to this underlying fragmentation. As Viera (2020) notes:

> ‘Temporal perception is initially fragmented. The perceived unity of time is constructed from these fragmented capacities. However, at present, no theory explains this unity. Something new is needed.’ (Viera, 2020, 655).

As I will show in Chapter 4, the mechanisms underlying time perception dissociate across at least five dimensions: (i) *modality* (vision, audition, etc.) (Burr et al., 2007; Fornaciai et al., 2018; Morrone et al., 2005); (ii) *temporal property type* (flicker rate, duration, etc.) (Eagleman, 2008b); (iii) *timescales* (sub- and supra-second) (Paton & Buonomano, 2018); (iv) *temporal judgment task* (Love et al., 2013; Mégevand et al., 2013); and (v) *temporal structure* (Arrighi et al., 2005; Nishida & Johnston, 2002, 2010).

To illustrate, consider the dimension of modality. It is possible to intervene in the visual processing of an audiovisual event in a way that affects our time perception of the visual information without similarly affecting our perception of the auditory information from the same event. Take the case of temporal ventriloquism described in the prelude, here a visual cue of an audiovisual event is perceived as occurring synchronously with an auditory cue, regardless of whether that auditory cue occurs just before or just after the visual cue because of the salience of the auditory cue (Hartcher-O’Brien & Alais, 2011).

A theory of temporal binding must connect these dissociable and fragmented mechanisms to the ways in which our temporal experiences appear unified. By partly relying on terminology from Viera (2020), I distinguish three relevant kinds of perceptual unity of time: temporal binding unity, timeline unity, and comparability. *Temporal binding unity* refers to how perceived features appear to be bound together in time relative to the moment we refer to as ‘now’. *Timeline unity* refers to how we experience the determinate temporal order of features from different modalities and distinct timescales by experiencing them as occurring within a single unified and coherent timeline that spans the recent past and the immediate expected future. *Comparability* refers to the phenomenological experience that the temporal relations of sensory features across modalities, timescales, and other dissociating dimensions belong to the same kind of representational property, such that comparing them seems effortless and automatic.

This contrast between our unified experience of time and the fragmented, dissociable mechanisms underlying temporal perception presents the third explanatory problem:

> ***Problem (iii):*** A theory of temporal binding must explicate the relationship between the mechanisms underlying temporal binding and the content and phenomenology of our temporal experiences.

In Chapter 4, I present this clash between our unified experience and the disunified structure of underlying mechanisms in detail. I also argue that the theory proposed in Chapter 3 provides a more satisfactory explanation of this relationship than existing accounts.

### *3.3.1 Potential Discrepancies Between Perceptual Phenomenology and Perceptual Capacity*

This third explanatory problem also crops up in another context, concerning methodological shortcomings in existing psychophysical studies. To gain a clear insight into the content and phenomenology of people's experiences of temporal binding, psychophysical studies must elicit both phenomenological reports about a participant's experience and results from tasks testing their capacity to engage in temporal binding. This allows one to assess whether there are discrepancies between reported phenomenology and our representational discriminatory capacities.

For example, a person may report that they can hear a sound and see a flash as occurring simultaneously, yet their response accuracy on this task indicates that they fail to correctly judge whether the features occurred simultaneously. If such discrepancies exist, they represent important data points that must be explained.

Unfortunately, psychophysical experiments often prioritise mechanical, quantifiable performance measures over detailed elicitation of subjective experience and lack protocols to carefully elicit detailed reports about the phenomenology of people's experiences (Haun et al., 2017). This leaves us with data that is limited and does not necessarily reveal everything we need to know about a person's experience of temporal binding in a specific psychophysical setting. If we do not have sufficiently sophisticated ways of gathering data on the content of people's experiences, but instead rely heavily on temporal tasks in which successful performance requires attention or other cognitive processing, we risk begging the question against certain theories of temporal binding that do not consider attention a necessary prerequisite for temporal binding (Arstila, 2016a, 2016b; Piper, 2019; Zeki, 2007) and against certain theories of consciousness that argue conscious experience can occur in the absence of attention (Aru & Bachmann, 2013; Block, 2007, 2019; Iwasaki, 1993; Koch & Tsuchiya, 2007; Lamme, 2006; Scholte et al., 2008; Tsuchiya & Koch, 2016; Zeki, 2015).

As such, even if it is true that temporal binding requires attention (as I argue in Chapter 3), or that attention is a prerequisite for conscious experience (as argued by several authors, see Baars, 2002; Dehaene et al., 2006; Dennett, 1991; Dołęga & Dewhurst, 2020; Naccache, 2018), we cannot presume that there are no discrepancies between temporal binding capacities and the phenomenology of temporal binding. It may be that people report that features in parts of their perceptual field appear as temporally bound even in situations where they could not possibly deploy attention to those features and therefore cannot represent this content and use it for discrimination tasks. Such reports must be explained because either they reveal important information about phenomenology that is not captured by the specified theory of temporal binding or they indicate that people tend to describe their experiences of temporal binding as richer than they actually are.

In Chapter 5, I provide a novel approach to overcoming the conceptual and theoretical issues involved in empirically investigating and interpreting the content and phenomenology of our experiences of temporal binding, and how these experiences relate to the mechanisms underlying temporal binding.

## 4 Upshot: Three Explanatory Problems for a Theory of Temporal Binding

I have argued that a theory of temporal binding must (i) account for the psychophysical data on how a wide variety of effects influence temporal binding, (ii) account for the problem of neural delay and the problem of desynchronisation, and (iii) explicate the relationship between the mechanisms underlying temporal binding and the content and phenomenology of our experiences of temporal binding.

These explanatory projects should not be treated as disconnected. If the mechanisms posited to account for how we engage in temporal binding render the content and/or phenomenology of our temporal experiences mysterious, then the theory introduces an explanatory gap between the workings of our perceptual system and the resulting experiences. Conversely, if we provide a comprehensive account of the content and phenomenology of our experiences of temporal binding without taking into account how the relevant mechanisms fit with psychophysical and neurophysiological data, the account should be rejected as empirically inadequate. Acknowledging the interdisciplinary and integrative nature of this project is essential.

In the next chapter, I analyse the brain time theory and argue that it falls short of addressing two of these three explanatory problems. In Chapter 3, I build upon existing event time theories to construct a new theory of temporal binding and show how it better addresses the first two explanatory problems: accounting for the psychophysical data and resolving the problem of neural delay and desynchronisation. In Chapter 4, I argue that this new theory also outperforms existing theories in addressing the third explanatory problem. In Chapter 5, I provide a novel approach to empirically investigate the claims of the theory presented in Chapter 3.

# Chapter 2

# Against the Brain Time Theory

## 1 Introduction

In this chapter, I argue against one popular kind of theory of temporal binding, the brain time theory (for supporters, see 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). As I presented in Chapter 1, the brain time theory essentially argues that the temporal order and synchrony relations that we perceptually experience are fully determined by the times at which the neural processes realise those experiences. These are determined by the perceptual endpoints of sensory features (Arstila, 2016a). So if the neural processing of an experience of a red colour and a green colour reaches their perceptual endpoint at the same time, we perceptually experience these two colours simultaneously and as occurring simultaneously. Importantly, what kind of neural event these perceptual endpoints refer to differs from theory to theory.

In this chapter, I argue that three different versions of the brain time theory, despite providing good neurophysiological models of some cases of temporal binding, fail to address Problem (i) which concerns accounting for the various psychophysical effects on temporal binding and Problem (ii) which concerns addressing the problem of neural delay and the problem of desynchronisation, briefly described in Chapter 1.

I start by setting out a general definition of the brain time theory in §2. In §2.1, I briefly contest the extent to which brain time theories are justified in claiming empirical support from a specific visual illusion that is paradigmatically taken to support the theory. In §3, I set out the problem of neural delay and the problem of desynchronisation (briefly mentioned in Chapter 1), in more detail. In §§4, 5, and 6, I discuss three different versions of the brain time theory that differ with respect to what kind of sensory processing they appeal to, to define perceptual endpoints. For each of these views, I argue why they cannot address Problems (i) and (ii).

In §7, I end by proposing that if we want to provide a theory of temporal binding that addresses Problems (i) and (ii), we should abandon the core tenets of the brain time theory and instead look to existing event time theories.

## 2 The Brain Time Theory

As I mentioned in Chapter 1, it is important to distinguish between the various uses of subjective time and brain time. I will therefore avoid referring to these vaguer terms and instead use the more precise vocabulary introduced in chapter 1. That is instead of subjective time I use either 'the time of representing' (i.e., the time a perceptual experience is consciously available) or 'the time represented' (i.e., the temporal content of our perceptual experience). Similarly, instead of brain time I refer to the 'onset of sensory signals' (the moment when a stimulus first impinges on a neural mechanism) or to 'perceptual endpoints' (the moment when a neural mechanism finishes its processing of some sensory feature and reaches the process responsible for conscious perceptual experience). In this chapter, we will mostly be concerned with perceptual endpoints and the different ways in which one can define such perceptual endpoints to formulate distinct brain time theories.

Because both consciousness and sensory processing are part of the definition of a perceptual endpoint, it is up for debate exactly what a perceptual endpoint is. Firstly, people may, based on their theory of what it takes for a brain state to satisfy the conditions of being a conscious perception, disagree about how to define perceptual endpoints. Perceptual endpoints could be defined as a local brain state depending on early sensory processing. One may take processing in some specific modularly defined cortex to be sufficient for conscious experiences of the locally processed feature (Zeki, 2007; Zeki & Bartels, 1999). By contrast, one could also define perceptual endpoints as a global brain state. That is, one may take conscious experience to require some higher-order processing, e.g., processing in the pre-frontal cortex, that integrates multiple sources of processing from across the brain (Dehaene et al., 2006, 207; Dehaene & Naccache, 2001).

Secondly, even if people agree that early intra-modal neural activity is sufficient for a brain state to satisfy the criteria for conscious perception, such that a perceptual endpoint can be defined as an intra-modal event, they might still disagree on what kind of processing this involves. That is, perceptual endpoints might be defined in terms of linear feedforward processing, such that features become conscious once they reach the end of some local linear neural processing activity (Zeki and Bartels, 1999; Zeki, 2007). Alternatively, perceptual endpoints may also be defined as involving both linear and non-linear processing, such that some features only become conscious when they reach the end of some non-linear processing activity (Arstila, 2015a, 2016a, 2016b).

Throughout this chapter, I will discuss three versions of the brain time theory, where these theories differ in terms of whether they define perceptual endpoints in terms of: (i) local linear processing (§4), (ii) local linear and non-linear processing (§5), or (iii) some global integrative processing (§6).

While the term 'brain time theory' is not always directly used, it denotes a popular theoretical framework embraced by cognitive scientists, psychologists, and philosophers (for early explicit and implicit supporters, see Arnold et al., 2001; Köhler, 1947; Zeki & Bartels, 1998, 1999). This view is still popular and has been repeatedly defended by Arstila (2015a, 2015b, 2016a, 2016b, 2019) and others working on time perception at the intersection between cognitive science, neuroscience, and psychology (Arnold, 2010; Moutoussis, 2012; Piper, 2019; Yarrow & Arnold, 2016; Zeki, 2007, 2015). In its simplest formulation, the theory states that brain time (understood as perceptual endpoints) determines subjective time (both the time represented and the time of representing). Different authors have put this relation differently:

> 'Experienced order in time is always structurally identical with a functional order in the sequence of correlated brain processes' (Köhler, 1947, 62).
>
> '[...] the brain does not necessarily bind together what happens in real time but may instead bind the results of the operations undertaken by its different processing systems, which require different amounts of time to complete their tasks' (Zeki & Bartels, 1998).
>
> '[...] the time course of perceptual experience correlates directly with that of neural activity' (Arnold, 2001, 596).
>
> '[...] the visual brain is characterized by different, functionally specialized systems, with different processing times for each; this type of organization naturally leads to different perceptual times for the corresponding attributes' (Moutoussis, 2012).
>
> '[...] the timing of our experiences is simply a matter of neural latencies. Each stimulus (or their features) are processed in parallel, and once the processing is completed—i.e., once the perceptual end point is reached—the stimuli (or their features) are experienced and they are experienced to occur in that very moment (e.g., not half a second ago)' (Arstila, 2016a, 168).

> 'I develop a detailed differential latency model of apparent motion to show how the ordinality of experiential content is isomorphic to the ordinality of relevant brain processes' (Piper 2019, 1).

As such, what features are temporally bound, and experienced as temporally bound, is determined by the times at which sensory information ascends through the processing hierarchy, reaching the requisite processes responsible for conscious experience. Arstila (2019, 340) states that, according to the brain time view, 'subjective time and brain time are isomorphic.' Using the more precise terms above, this means that the distinction between 'the time represented', 'the time of representation', and 'perceptual endpoints' is a mere conceptual one. As such, the brain time theory presents a very clear empirical thesis. Namely, that tracking the neural activity of when sensory processing finishes processing (however one defines this) is also akin to tracking how we represent sensory features as being temporally related. This is captured by the brain time theory's endorsement of the following thesis:

> ***The thesis of temporal isomorphism:*** '[T]he contents of our experiences and the neural states that underlie them share the same temporal properties' (Arstila, 2016a, 167).

As Arstila (2016a, 168) puts it: '[...] if the thesis [of temporal isomorphism] is correct, the order in which stimuli are experienced to occur is the same in which the processing related to contents is completed.' The thesis of temporal isomorphism is paired with another thesis called the thesis of minimal delay (Arstila, 2016a):

> ***The thesis of minimal delay:*** 'our experiences of external events are only delayed by the time it takes for light and sound to reach our sensory receptors and for our neural mechanisms to process the stimuli.' (Arstila, 2016a, 164).

Any minimal delay in the unconscious local processing of sensory features entails a minimal delay in when we consciously perceive those features. No delays are added in the brain to compensate for the differential processing times of our sensory mechanisms.

The thesis of temporal isomorphism and the thesis of minimal delay make up the core of brain time theory. As such, any theory of temporal binding that rejects either of these theses cannot reasonably be considered a brain time theory. For example, a theory that takes temporal binding to rely on processes or principles that temporally organises our representations of sensory features in a way that clashes with the order and timing of perceptual endpoints, is not a brain time theory (as this violates temporal isomorphism). Similarly, a theory that allows for some kind of temporal compensation, such that the time at which a sensory feature is represented as occurring precedes or is delayed relative to the otherwise ordinary timing of the perceptual endpoint, is not a brain time theory (as this violates minimal delay). This should serve as a guide for what it takes for a theory to be a brain time theory, rather than some other theory of temporal binding.

By using this definition, it should be clear how to assess whether the empirical evidence supports or counts against brain time theories. The two core theses allow a brain time theory to make precise empirical and clear predictions of how we engage in temporal binding. It can in principle, predict the time represented of any given perceptual experience on the basis of information about the exact timing of the perceptual endpoints of the sensory mechanisms in play. This simplicity and empirical predictive power is an important aspect of the theory (Arstila, 2015b; Moutoussis, 2012; Yarrow & Arnold, 2016). Therefore, if it turns out that we can accurately predict the content of people's experiences of synchrony and temporal order from information about the timings of perceptual endpoints, this will count as support for the theory. However, this also means that if the relevant facts about the neural processing times and delays involved in a perceptual experience do not predict the time represented, then this counts against the theory.

In §4, §5, and §6, I argue that there are several studies on how we temporally bind sensory information that these different versions of the brain time theory cannot account for. Before I discuss these cases, I want to briefly look at one kind of empirical study that has been at the centre of discussions concerning the empirical support for the brain time theory.

### 2.1 An Aside on the Colour-Motion Asynchrony Illusion

Advocates of the brain time theory have repeatedly taken the colour–motion asynchrony illusion to empirically vindicate the theory (Moutoussis & Zeki, 1997). In this illusion, participants see a red pattern moving leftward, which then switches to a green pattern moving rightward at some set rate of change (see Figure 1). When this pattern oscillates quickly (1–2 Hz), people cannot confidently pair colour and motion direction and tend to report that colour changes before motion (to

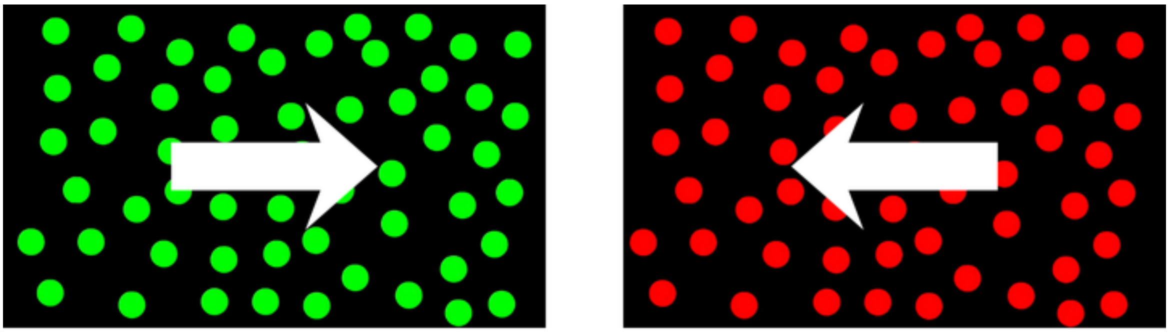

**Figure 1:** Pattern of circles (sometimes squares) randomly scattered across the entire screen oscillating between being red or green and between moving either left or right. The task is to pair which colour pattern goes with which motion direction.

experience this animation at different speeds click **here** and see Animation 1a, 1b, 1c). However, if the colour change is delayed by approximately 80–100 ms, most errors in temporally binding the two features disappear (Moutoussis & Zeki, 1997).

This has been taken to support the view that differential processing times of colour features and motion features explain the apparent asynchrony. Arnold et al. (2001), Arnold (2010), Moutoussis (2012), Zeki (2007), and Zeki and Bartels (1998) argue that the reason why synchronous changes in motion and colour are perceived to occur asynchronously is that colour is processed faster than motion. So, when the colour change is artificially delayed relative to the change of the motion direction, the neural processing of the motion direction change and colour change reach their perceptual endpoints at the same time, causing us to perceive these as bound. See for example Arnold's interpretation:

> 'The similarity between our data and earlier findings regarding the asynchrony of visual consciousness [Moutoussis & Zeki, 1997; Zeki & Bartels, 1998] strongly suggests that the asynchrony of visual consciousness is a direct consequence of asynchronous neural processing.' (Arnold, 2001, 600)

However, it is unclear how much credence should be lent to this interpretation and thus how much these studies support the brain time theory. The neural processing asynchrony interpretation has been challenged by Nishida & Johnston (2002, 2010). According to them, differences in feature processing times do not explain the perceived asynchrony. Rather, it is differences in the temporal structures of colour and motion features in the illusion that explain the apparent asynchrony. The colour change is a 'first-order change', insofar as it is a static derivative that is defined over two successive points (locations) in time (Nishida & Johnston, 2010). By contrast, the motion direction change is a 'second-order change', insofar as it is a 'second-order temporal derivative, whose definition requires at least three successive points in time' (Nishida & Johnston, 2010, 287).

Nishida & Johnston suggest that there are substantial differences in how we discriminate first-order changes (e.g., sudden changes in an attribute like colour or position) and second-order changes (e.g., motion reversals of a continuously moving object). Evidence suggests that first-order changes produce a salient transient that can be picked up by fast early-level sensory mechanisms that track the onset of these transients (Nishida & Johnston, 2010; Werkhoven et al., 1993). Second-order changes, however, do not produce salient transients but rather produce temporally uncertain cues. The perception of second-order changes seems to depend on slow mid-level mechanisms, which must compare at least three successive points in time (instead of just two), e.g., it must compare an initial location of a moving stimulus to a later location (to determine direction), and then compare to a third later transient to determine whether the stimulus continues along its trajectory or changes direction (Nishida & Johnston, 2010; Werkhoven et al., 1993). This also means that second-order changes produce no unique clear transients and lead to temporally uncertain discriminations.

This is supported by Nishida and Johnston (2002) who show that second-order changes appear to be delayed relative to first-order changes in an irregular manner, regardless of what kind of features occupy these kinds of changes. Arstila (2015a) concedes that this might falsify the claim that colour features reach their perceptual endpoint before motion features, thus being incompatible with certain brain time theorists like that of Arnold (2001), Moutoussis (2012), and Zeki (2007) (discussed in §4). However, it may still be compatible with some versions of the brain time theory. If for example, 'the mechanism responsible for temporal judgments in the brain time view requires more time to determine turning points [second-order changes] than to determine transitions [first-order changes]' (Arstila, 2015a, 3)

A different study on these colour-motion pairings, however, puts additional pressure on the brain time theory's interpretation of the colour-motion asynchrony illusion. Holcombe and Cavanagh (2008) show how exogenously probing attention to the transients[8] produced by the colour and motion with an extrinsic transient cue (a superimposed bright red ring) can reduce or even eliminate the otherwise apparent asynchrony of the colour and motion. Instead of delaying the occurrence of the colour change relative to the motion change to remove the apparent asynchrony (as in Zeki and Bartels, 1998), they only manipulate attention.

Holcombe and Cavanagh (2008) argue that this is because the extrinsic transient cue probing attention gives the perceptual system equal rapid access to the transient of both colour and motion at the time of the cue. These results cannot be predicted by the brain time theory. If the time represented is determined by the time at which features finish processing at some local perceptual endpoint of sensory signals (be they first-order or second-order changes), then the relative feature timing would be settled independently of how attention later accesses the representational content of distinct signals. Even if attention affects processing times directly, 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, which it is not (Holcombe & Cavanagh, 2008, 14).

---

[8] The rapid sensory signals that can be picked up automatically by spatial attention.

There are thus plenty of reasons to doubt whether the colour-motion asynchrony illusion, frequently cited as evidence for the brain time theory, in fact is accounted for by the brain time theory. It is at least clear that this colour-motion asynchrony illusion does not uniquely support the brain time theory, and it even has difficulties accounting for variations on it.

In this chapter, however, I want to expand the scrutiny of the brain time theory beyond this case to discuss how the brain time theory might deal with accounting for temporal binding more broadly. That is, whether it can satisfy Problems (i) and (ii) presented in Chapter 1. In the next section, I set out Problem (ii) in more detail. In §§4, 5, and 6, I then discuss whether or not the different versions of the brain time theory can account for Problems (i) and (ii).

## 3 The Problem of Neural Delay and the Problem of Desynchronisation

The problem of neural delay and the problem of desynchronisation arise from the fact that sensory processing takes time and that the processing of different features takes different amounts of time. I describe each of these problems in turn.

While the processing of sensory features takes time, our perception of multi-modal events appears to be consciously available more or less at the same time as they unfold in event time (Eagleman, 2008; Hogendoorn, 2022; Johnson et al., 2023; Nijhawan, 2008).

Holcombe (2009) emphasises two very different limits on the speed of visual processing. There is low-level processing, which operates at a high temporal resolution in a mostly feedforward manner, and higher-level processing, which operates at a much lower temporal resolution and relies on later integrative processing as well as attentional mechanisms. Low-level processing is important for the detection of first-order changes in colour, contrast, orientation, etc. (see Holcombe (2009)). The higher-level processing is important for the detection of individual objects consisting of multiple features and how these change relative to each other over time. This includes binding local features into global forms, binding global forms with colour, and perceiving higher-order motion (Holcombe, 2009).

Low-level processing has a high temporal resolution (approximately as high as 30 Hz) and begins very quickly after stimulus onset. Within the visual modality, it takes between ~30–70 ms (after stimulus onset) for retinal information to activate the earliest cortical responses for visual processing in V1 (Maunsell & Gibson, 1992). This delay is due to the time required for photoreceptors to detect the relevant stimulus and for the produced signal to propagate through ganglion cells until it reaches V1, where it undergoes early sensory processing. For auditory processing, it takes only ~10 ms to activate the earliest cortical responses (Pratt, 2011). Higher-order visual processing has a lower temporal resolution (in some cases at most 3 Hz, in others at most 9 Hz), and begins at around 100 ms after stimulus onset and develops across multiple stages (Holcombe, 2009). White (2018) states that 'object identification [involving semantic processing] probably begins around 150–200 ms' after stimulus onset. This highlights the potentially slow and multi-stage nature of visual processing, at least when the perceptual targets in question are complex multi-featural objects and events.

But if our perceptual experience of objects and events lags behind event time by such a substantial amount of time, why do we not notice this, and why are such delays not detrimental to the coordination of a range of sensorimotor tasks that are guided by conscious perception? Consider how a relatively slow object like a normal cyclist (travelling 25 kph) would travel a meter (1.04 m) in just 150 ms.

There is, as such, an explanatory problem of accounting for how our perceptual system overcomes delays in neural processing time to represent complex multimodal objects in the world in time to perceptually guide time-sensitive actions and present us with perceptions that make it appear that we represent the world in real time. This is the problem of neural delay.

The second problem, the problem of desynchronisation, refers to the problem of accounting for how we temporally bind sensory features into coherent perceptions such that we accurately perceive the synchrony and temporal order relations of those features, even when the processing of these features takes different amounts of time. While our perceptions are multimodal, that is, integrated perceptual experiences of multiple sensory features and modalities, the processing of these sensory features desynchronises at several stages between the time at which they occur in the environment and the time at which they finish processing.

First, sensory signals desynchronise because they propagate at different speeds. For instance, light travels faster than sound, meaning visual and auditory information will arrive at our sensory receptors at different times. This temporal discrepancy is compounded by the fact that these signals hit different sensory receptors at varying moments. For example, sound waves may enter the right and left ears at slightly different times, and a touch on the toe has to travel further than a touch on the cheek. All of this creates a temporal asynchrony between signals that the brain must reconcile.

Second, the processing of sensory information occurs in parallel across modalities, but the temporal dynamics of this processing vary considerably across sensory features (Bullier, 2001; Nishida & Johnston, 2010; Pöppel, 1988; Zeki, 2015). Neurophysiological studies of the visual cortex in macaque monkeys, which function as a model of human sensory processing, have shown that latencies (the delay between stimulus onset and the neural response) vary across different visual areas: 40–100 ms in V1 (early spatial organization), 50–100 ms in MT (motion processing), 70–160 ms in V4 (surface properties, shape, motion features), and 90–180 ms in IT (visual recognition) (Bullier, 2001). These discrepant processing speeds create a coordination problem in sensory processing within modalities as well as between modalities.

Third, the neural processing of different features (originating from the same event) is not only temporally desynchronised but also involves complex interactions, even within one modality. As mentioned above, the processing of sensory features such as colour and shape can interfere with one another, such that the processing times of certain sensory features are dramatically altered if they are to be bound with each other (Holcombe, 2009). We can perceive the first-order motion of bright peaks of a black/white grating through a window even when the white bars pass each perceptible location 30 times per second. However, when we have to perceive the changes of a pattern that changes colours and shape simultaneously, we can only perceive the shape-colour pairing when the changes are very slow (below 2–3 changes per second). Additionally, intermodal interactions add another layer of complexity. Information from one modality can influence how we process sensory data from another. For example, tactile sensory processing has been shown to influence early visual processing (Groen et al., 2013). This suggests that neural processing in one modality can resynchronise, or potentially disrupt, the perceived synchrony of sensory features from other modalities, making the process of temporally integrating different sensory features even more challenging.

In effect, the features originating from one event desynchronise as they propagate, hit our peripheral nervous system, are processed in parallel, and pass up and down the processing hierarchy. There is, as such, an explanatory problem of accounting for how our perceptual system manages to temporally bind features, such that we can perceptually experience multimodal events as temporally bound given such desynchronisation. This is the problem of desynchronisation.

A theory of temporal binding must give a principled and coherent answer to how our perceptual system organises information in the face of neural delays and desynchronisation.

In the following sections (§§4, 5, and 6), I critically discuss whether the different versions of the brain time theory can account for the problem of neural delay and desynchronisation (Problem (ii)). In doing so, I also present a range of empirical studies on psychophysical effects on temporal binding that are difficult, if not impossible, to explain by reference to the resources available to the brain time theory (Problem (i)). I start by discussing a version of the brain time theory, according to which perceptual endpoints are defined in terms of local feedforward processing.

## 4 Perceptual Endpoints as Local Linear Processing

One understanding of perceptual endpoints comes from Zeki and others (Zeki 2007, 2015; Zeki & Bartels, 1998; Moutoussis & Zeki, 1997), who argue that perceptual endpoints can be defined in terms of the moment when a stimulus (e.g., colour or motion) has been processed in its respective cortical area (e.g., V4 or V5, as in the case of colour or motion).

This means that when a stimulus (such as colour) finishes processing and becomes consciously available (as a perceptual experience of colour), it relies on a 'straightforward computational process that engages the colour centre and the areas feeding it.' As Zeki puts it:

> 'Processing in an area may have a conscious correlate (i.e., become perceptually conscious and not require further processing) (Zeki & Bartels, 1999), without necessarily depending upon some kind of 'top-down' influence from higher areas.' (Zeki, 2007, 582-583).

Zeki takes this kind of linear processing to straightforwardly determine the perceived temporal order of stimuli (as mentioned in §2.1). In the next two sections, I discuss how a brain time theory with this conception of perceptual endpoints fails to account for Problems (i) and (ii).

### 4.1 Accounting for the Problem of Neural Delay

According to the thesis of temporal isomorphism, a feature is experienced the moment it reaches its perceptual endpoint, i.e., the time represented covaries with the timing of perceptual endpoints. This also means that every bit of processing delay in the processing of some specific feature adds delays not only to the time of representing that feature, but also to the time represented (Nishida & Johnston, 2010). Adherents of this version of the brain time theory, such as Moutoussis (2012), are of course not oblivious to this. They might just think, as Arstila (2015a) has argued elsewhere, that this need not be an issue if these delays are minimal, as suggested by the thesis of minimal delay.

However, the necessary delays for transmitting and processing are not always minimal enough to account for all cases of how our brain engages in temporal binding. That is, the neural delays involved in direct processing of stimuli may not match up with experimental evidence of the times at which we perceive sensory features as occurring. If the brain time theory is right, then the neural delays involved in direct processing of sensory features may mean that perceptually we live so much in the past that it would be unable to explain how we engage in many kinds of semi-fast ball games and navigate high-speed traffic conditions.

One way of accounting for differences in our fast action and our sluggish perceptions of the action and events themselves is to appeal to the distinction between the dorsal and ventral route of visual processing. The dorsal route is argued to be a fast, specialised processing route (~100 ms), associated with spatial information such as motion and depth. The ventral route is a slower processing system (~150–200 ms) associated with the processing of features such as shapes, object recognition, and other properties, allowing for rich perceptual representation of objects (Goodale et al., 2004). The brain time theorist could use this to argue that while the dorsal system furnishes fast perceptual responses that are used to guide action, phenomenal visual perception is a separate and delayed process undertaken in the ventral processing stream.

There is some merit to this answer, and one could argue that these two kinds of visual processing are largely asynchronous and fragmented (as does Zeki, 2015; Zeki & Bartels, 2009). While the dorsal stream supplies quickly accessible information guiding our actions, information from the ventral stream synchronises with this dorsal stream in order to gradually develop an integrated visual and phenomenal perception (processing in the two streams is only ever about 100 ms apart, so synchronisation would not be much delayed). This integrated perception can, in turn, be used to guide more complex actions that do not rely solely on the fast dorsal processing of motion and depth information. On this view, visually guided actions that first rely on quickly processed motion and depth can later be accompanied with rich conscious visual perceptions that develop later and across time as information is integrated across distinct visual pathways (see also Hollingworth, 2008; Zimmermann et al., 2014).

One reason why this answer is unsatisfactory is that we do not always seem to experience the kind of perception–action asynchrony (White, 2018) that would be generated by treating dorsal and ventral visual processing as fully dissociable systems (something that is also itself empirically implausible, see Goodale et al., 2004). Consider the case of a game of tennis. A professional tennis player can hit the ball so that it flies well over 350 km/h, i.e., 9 metres in 100 ms. However, if the visual perceptual processing of the position of the ball takes between 150–200 ms, then the perceptual endpoint of the phenomenal visual (ventral) processing of the position of the ball would be delayed so much that one might return the opponent’s ball before seeing them hit it.

Dissociating the ventral and dorsal stream in this way would create a large asynchrony between our perceptions and our actions (an asynchrony that we do not seem to perceive). When we play tennis, there do not seem to be such significant delays between our action of hitting the ball and our perception of hitting and hearing the ball being struck. Moreover, a tennis player can rapidly shift their gaze to the location of the ball. Gaze direction does not lag behind the actual location of the ball (see White, 2018, who uses Badminton as an example). If the tennis player actually has a conscious visual perception of the ball at the location of their gaze, then that would be impossible to explain (because of neural delays in the ventral system). The most recent available perceptual information about the ball's location would be located several feet away from the direction of one's gaze. So, if we can engage in gaze following of fast-moving objects and perceive the ball at that position, which we seem to do, then this cannot be explained by the brain time theory.

This problem is further compounded by the fact that there is another available explanation: our perceptual system may engage in predictive extrapolation to perceive the trajectory of objects and mitigate delays incurred by neural processing times. This may cause us to have a predictive perceptual representation of a moving object, such that the time of representation of the object is closer to the actual event itself, than if we had to rely on direct processing (and the delays involved in the ventral stream). A series of cases by Blom et al. (2020, 2021), Johnson et al. (2023), and Turner et al. (2024) show how our perceptual system actually pre-activates motion processing to process motion features at locations ahead of where the moving stimulus actually is. This kind of extrapolation allows our perceptual system to engage in cortical processing of a non-existent stimulus based on predictions about its likely location, thus enabling us to perceive motion and trajectories in close to real time.

But how can the brain time theory explain the fact that we perceive the motion trajectory of objects almost in real time, when the perceptual endpoints of the signals emitted by the moving object would lag behind reality? According to the brain time theory, the content of reports is isomorphic to the timing of the perceptual endpoints. However, if perceptual endpoints lag behind the time represented by predictive models then we cannot explain the time represented by appealing to perceptual endpoints. This indicates a general problem with the idea that the time represented is fully determined by local linear perceptual endpoints.

One might attempt to defend the brain time theory by appealing to predictions in the specific cases where the time of representing seems to precede the perceptual endpoints of feedforward processing. However, these kinds of predictive perceptions might not be compatible with the brain time theory.

Given the thesis of temporal isomorphism, we do not represent features before, but exactly when they reach their perceptual endpoints. However, since the time represented through perceptual predictions precedes perceptual endpoints, the brain time theory would need substantial reformulation to allow for predictive perceptions. Predictive perceptual processing may give rise to a special type of perceptual endpoint that does not depend on local, parallel feedforward processing, but instead depends on ingrained predictive models that exert top-down influence on the time represented in our perceptions (as in predictive processing models by Clark, 2013; and Hohwy, 2012).

But how do such predictive models fit within a brain time theory? In other words, what are the perceptual endpoints of predictive perceptions oriented toward? They cannot be oriented toward direct processing times, because direct processing is bypassed in these situations. Therefore, the perceptual endpoints of predictions must be derived from the expected event time of external sensory features.

In this case, however, the perceptual endpoints of predictive perceptual processing are not motivated by brain time theoretical considerations but rather by event time theoretical considerations. Remember the event time theory argues that our perceptual systems have mechanisms in place that compensate for neural processing times to approximate veridical event time perception. Predictions of how external events are temporally related are geared towards overcoming neural processing times and depend on top-down expectations to approximate veridical event time perception. Therefore, it is no longer clear, in cases of predictions that the time of representing and the time represented are isomorphic. As such predictions clash with the brain time theory's thesis of temporal isomorphism. Moreover, as discussed in §2, a key motivation for the brain time theory is that it allows simple and elegant prediction of the time represented as a consequence of the neural processing times of local feedforward processing (Yarrow & Arnold, 2016). If perceptual endpoints are malleable depending on context-dependent top-down predictions, this motivation is undermined.

Therefore, although the brain time theory seems to have the tools to account for how the brain deals with the problem of neural delay by arguing that perceptions of features are available the moment they are linearly processed, it is, as formulated here, unable to account for how we perceive action–perception synchrony. It also cannot account for cases where conscious perceptions of temporal relations are available faster than direct feedforward processing allows. Although many such cases exist, this might not be fatal to the brain time theory. The theory could potentially be reworked to include predictive processing, which might adjust for the neural delays of our sensory processing system based on statistical regularities in the environment. However, it is not clear what such an account, consistent with the core principles of brain time theory, would look like. We now turn to the problem of desynchronisation.

### 4.2 Accounting for the Problem of Desynchronisation

The problem of desynchronisation concerns explaining how our perceptual system can temporally bind sensory information across several modalities, allowing us to perceive the synchrony and temporal order of external features and events, when the processing of these features is desynchronised at several stages relative to their occurrence in the external world.

According to the version of the brain time theory under consideration, whether two sensory features are perceived as asynchronous (occurring in some temporal order) or synchronous depends solely on the timing of perceptual endpoints, and not on whether the features occur at the same time in event time. Moreover, due to the thesis of minimal delay, it is hypothesised that there are no compensation mechanisms in place, such that certain perceptual endpoints are temporally postponed in order to align perceptual endpoints of features originating from the same event.

However, as mentioned in §3, the processing of different sensory features varies. If this version of the brain time theory is correct, we would expect to perceive simultaneous changes in the different features of a multimodal event as occurring in some temporal order dictated by local linear perceptual endpoints, rather than as occurring synchronously. Accepting the facts about differential neural processing times and the two core theses of brain time theory, it follows that it would be practically impossible for our perceptual system to accurately temporally bind complex multimodal events. Our perceptual system would thus be a massively asynchronous one, with each feature's perceptual endpoint making up a micro-conscious experience, as is in fact suggested by Zeki (2015).

This is problematic, given that multiple psychophysical studies indicate that temporal perception often compensates for the differential processing times of sensory signals. Alais & Carlile (2005) demonstrate how the brain compensates for the slow speed of sound relative to light. In their experiment, participants were presented with auditory and visual stimuli, with the auditory source's distance manipulated. Simply altering the apparent distance of the auditory signal (through a strong auditory-depth cue) affected the temporal binding of the visual and auditory stimulus. When the perceived auditory source was simulated as being further away, a longer delay needed to be inserted for participants to perceive the two stimuli as simultaneous.

This suggests that temporal binding is influenced not only by differential processing times, but also by contextual cues that are suggestive of a specific temporal relation.

Other studies show similar effects. The perceptual system compensates for potential asynchronies caused by spatial distance and neural processing latencies, generating perceptual experiences of the world that match the synchrony and temporal order of features as they occur in event time (see Chen & Vroomen (2013) and Vroomen & Keetels (2010)). As presented in the prelude, Hartcher-O'Brien and Alais (2011) demonstrate a temporal version of the ventriloquism effect, in which two temporally separated cues (a sound and a flash) are perceived as bound based on the timing of the more salient cue, regardless of presentation order. This illustrates that temporal binding depends not only on neural delays but also on the saliency and context of the stimuli.

Many other cases support a similar dissociation between the time represented and neural processing times. For example, causal expectancies (Bechlivanidis et al., 2022; Haggard et al., 2002; Stetson et al., 2006; Stone et al., 2001), adaptations to repetition (Di Luca et al., 2009; Fujisaki et al., 2004; Vroomen et al., 2004), and exogenous attentional cues (Chennu et al., 2009; Holcombe & Cavanagh, 2008; McDonald et al., 2005) all influence the time represented without influencing neural processing times.

Even in the simple case of colour–motion feature pairing (see §2.1), often cited in support of brain time theory, the problem of desynchronisation becomes evident. If colour is truly processed faster than orientation (as claimed by Arnold (2001), Zeki & Bartels (1998), and Zeki (2007)), then as White (2018) argues, information about the orientation of a projectile moving at 90 mph would lag approximately 8 feet behind information about its colour. Yet we do not perceive these features as lagging across our retina. We see a single, sharply defined object moving across our visual field. As is also noted by Holcombe (2014), such perceptions are often more sharply defined than the temporal resolution of visual processing would suggest. White (2018, 645) states that this is not just a problem of feature binding, but also of synchronisation, since ‘asynchronous outputs have to be bound into a synchronous representation of a moving object where all of the features are in the same location in visual space at a given time,’.

As such, the delays and timings involved in the linear processing of sensory signals do not provide the full story of how our brain engages in temporal binding. Consequently, this version of the brain time theory lacks a satisfactory answer to the problem of desynchronisation.

## 5 Perceptual Endpoints as Local Non-linear Processing

The brain time theory struggles to account for desynchronisation because it ties temporal binding to perceptual endpoints. However, this issue might arise only if we define perceptual endpoints in terms of linear, local feedforward processing. In this section, I consider non-linear processing views (Arstila, 2015a, 2015b, 2016a, 2016b; Lamme, 2006; Lamme & Roelfsema, 2000; Piper, 2019) that challenge this assumption, arguing that sensory processing may also involve interactions between higher and lower cortical areas before it can be considered as having reached its perceptual endpoint.

By taking Arstila's (2015b, 2016a) 'non-linear latency difference view' as an example of non-linear processing views, one can account for certain effects of retrodiction. These are cases where a later signal at t2 influences the perceptual interpretation of an earlier signal at t1. Such retrodictive effects would otherwise present a problem for the brain time theory, since theories based on feedforward processing alone cannot explain how a stimulus processed later can affect the perceptual interpretation of a stimulus processed earlier.

Arstila's (2016a) non-linear latency difference view is inspired by models of re-entrant processing (Lamme, 2006). Re-entrant processing occurs when the processing of a signal in higher-level cortical areas reactivates processing in an early lower-level cortex. Re-entrant activity facilitates an exchange of information between systems by allowing information to pass forward and backward between different sensory mechanisms at many different times. This allows for the perceptual binding of processed sensory features that are otherwise not processed together (Engel & Singer, 2001). Re-entrant processing may cause a perceptual endpoint to take longer to be reached, as the processing involves the integration of information transmitted between various cortical areas. In such cases, the perceptual endpoint lies at the 'end' of a non-linear processing activity (Lamme, 2006).

These re-entrant perceptual endpoints can be understood in terms of either local or global re-entrant activity. Local re-entrant perceptual endpoints are defined in terms of re-entrant activity that originates within the cortical area specialised for processing the feature in question. For example, within the visual cortex, re-entrant processing might involve a signal being processed in V5, after which it re-enters and terminates in a lower-level visual processing area, such as V1. People who advocate for local perceptual endpoints would likely side with those who argue that recurrent early sensory processing in, for example, the occipital lobe, is sufficient for a brain state to count as a conscious perceptual state (Block, 1995, 2007, 2019; Iwasaki, 1993; Koch & Tsuchiya, 2007; Lamme, 2006; Lamme, 2003).

By contrast, global perceptual endpoints are defined in terms of re-entrant activity originating from higher-order cortices, such as the prefrontal and parietal cortex, which are correlated with cognitive processes such as selective attention, thinking, reasoning, reporting, decision-making, memory functions, etc. (Block, 2019). On this account, a signal processed locally in V1 may have to be modulated by some higher-order cortex, which makes V1 accessible to other cortices also modulated by higher-order cortical activity, in order to become consciously available as part of some integrated conscious perception. Those who would subscribe to global perceptual endpoints would likely take sides with those who argue that prefrontal activity is necessary for a brain state to count as a conscious perceptual one (see for example Brigard & Prinz, 2010; Brown et al., 2019; Dehaene et al., 2006).

The brain time theory discussed in this section assumes that local re-entrant processing is sufficient for conscious perceptual experiences (Arstila 2016a), and thus that local perceptual endpoints exist. In §6, I present a distinct version of the brain time theory that relies on global perceptual endpoints.

If we take re-entrant processing of some specific sensory feature (e.g., colour) to determine the local perceptual endpoint of that feature, then Arstila's account allows for cases of non-linear processing. For example, the perceptual endpoint of some sensory feature S1 may refer not only to its feedforward processing in V1 (early visual processing) and processing in the later cortical layer of V4 (early visual colour processing) but also to its subsequent *re-entrant* processing in V1 where it is processed again before it can be said to reach its perceptual endpoint.

With the addition of non-linear processing, perceptual endpoints are not necessarily fixed by a single feedforward sweep. Instead, they can also be the result of re-entrant activity that integrates sensory signals across multiple stages of processing. This more flexible notion means that features that may be desynchronised in their feedforward processing can nevertheless fuse to form a unified perceptual experience if they reach their non-linear local perceptual endpoint at the same time. In effect, by relying on re-entrant loops, the perceptual system can perceptually integrate features that occur at different times in the external world.

For example, a colour feature occurring at event time t1 might be processed in V1 at t2, proceed to V4 at t3, and reach its perceptual endpoint by re-entering V1 at t4. Moreover, a second feature (a colour feature) occurring at event time t3 might reach its perceptual endpoint in V1 at t4. In this way, two sensory features (e.g., motion and colour) occurring at different event times, t1 and t3, might be processed together in an integrative fashion at V1 at t4, reach their perceptual endpoint simultaneously, and be experienced as occurring synchronously. This non-linear processing can thus be used to account for certain aspects of the problem of desynchronisation, which will be the focus of this section.

Note, however, that the non-linear processing account presented here does not directly add anything of value in accounting for the problem of neural delay as discussed in §4.1. That is, it does not account for the lack of perception-action asynchrony or explain how we may rely on perceptual predictions when engaged in temporal binding of object trajectories. The non-linear processing account allows for top-down influences on sensory processing (through re-entrant processing), and this could perhaps be used to add some kind of predictive processing story to the account. However, this has neither been done nor is it clear how well such an account aligns with the conceptual framework of the brain time theory (for reasons given in §4.1). Thus in this section I only discuss how the non-linear processing account might better deal with the problem of desynchronisation than the feedforward processing account.

Arstila (2016a) uses his non-linear processing account to explain the retrodictive effect of the colour-phi motion illusion. In this illusion, two discs, a red disc 'A' and a blue disc 'B' are flashed side by side in quick succession. This sequence of stimuli is perceived as a continuous motion of disc A moving to the position of disc B, with the colour changing mid-trajectory. The puzzle of this phenomenon is to explain how a person can perceive motion and colour changes during the gap between the stimuli before perceiving the second stimulus at its actual position. The only possible explanation, given that we assume no violation of causality, seems to be that the time represented and time of representing come apart (violating temporal isomorphism) through some sort of backward influence between temporally separated processing of sensory features. So the question is, how does brain time theory explain this illusion of apparent motion without violating either the physical laws of causality or its own thesis of temporal isomorphism?

Arstila (2016a) gives the following explanation. Although most sensory information from the retina first passes through an area of early visual processing (V1), some sensory information of the second disc may bypass this early processing in V1 altogether and directly activate V5, which is responsible for processing motion (Sincich et al. (2004) present evidence that visual input can bypass V1 to activate V5). This, in turn, can be used to determine the retinotopic location of the second disc. In the case of illusory motion, the appearance of the retinotopic location of the first and second discs could thus activate processing in V5 with or even before the processing of the shape and colour of the second disc has begun in V1. The superposition of the re-entrant processing of the first disc in V1 and the feedforward processing of the second disc, bypassing V1 to V5, is thus responsible for apparent motion perception.

Arstila takes this to have the consequence that the perception resulting from re-entrant processing ‘does not need to be specific about its cause [the event that caused it]’ because it is simply the result of a fused perception due to some set of perceptual endpoints being synchronous (Arstila, 2016a, 180).

This explanation fits with the principle of isomorphism, as the sensory features of the apparent motion and the second stimulus are experienced when their re-entrant processing terminates in V1 (at the end of some non-linear processing sequence). Arstila concludes that the retrodictive illusions which otherwise seem to require some delay or recalibration mechanisms can be explained as the consequence of non-linear processing (for a similar argument using a different model, see Piper 2019, 5-9).

While this is a promising account, I argue that extending the brain time theory to include non-linear processing as well as linear processing suffers from three issues concerning accounting for the problem of desynchronisation more broadly, which make it implausible as a general account of temporal binding (despite being a promising way to model various retrodictive phenomena, see Arstila, 2015a, 2016a, 2016b).

First, in ecological settings, different sensory features (such as motion and colour, or auditory and visual signals) are processed in parallel via distinct neural pathways with different latencies. If perception were merely a matter of tracking the physical arrival times of stimuli, then events with different processing times should be perceived as temporally misaligned. However, this is not always what happens. Often, we perceive multimodal events as occurring in synchrony, even when the neural processing times of the features that make up the event differ significantly (see §4.2). This raises the question: How does the brain achieve this perceptual synchronisation? While the addition of non-linear processing allows for some flexibility in perception, as seen in retrodictive effects like the colour-phi illusion, it breaks down when it comes to accounting for how we integrate multiple simultaneously occurring stimuli (e.g., auditory and visual components of speech).

If perceptual endpoints rely on local re-entrant processing and multiple stimuli re-enter earlier cortices at different rates, aligning perceptual endpoints relative to event time would be extremely cumbersome. This is because re-entrant signals from higher cortical areas arrive in early processing areas (such as V1) at the same time that new incoming sensory information is arriving from the external world. This results in a mixture of bottom-up and top-down information within the same neural space, making it unclear how the system prevents cross-contamination between sensory information originating from different events.

For the general coordination of top-down and bottom-up signals, such a system would require a meta-perceptual function capable of aligning higher-level cortical re-entrant signals with linear feedforward signals. This would necessitate a complex system that categorises incoming signals into (a) feedforward signals and (b) re-entrant signals. To correctly determine event timing, these signals would need to be aligned relative to the event from which they originate. This would ensure that the correct non-linear top-down and linear bottom-up signals align. However, while this may work in some cases, it is an unrealistic and biologically implausible mechanism for ensuring precise coordination between top-down and bottom-up processing.

Without such a system in place, there is no way to ensure that the relevant feedforward and re-entrant signals are correctly integrated (avoiding cross-contamination). Moreover, for such a mechanism to work, the brain would need a pre-existing model of external time to guide coordination. This would require something like the predictive processing system proposed in §4.1, where perception is shaped not just by sensory input but also by prior expectations about how events unfold. However, if temporal perception depends on expectancies rather than the timing of neural processing alone, the brain time theory would need to be reframed. Instead of explaining temporal binding as a function of processing delays, it would have to account for the brain's active role in structuring time perception based on predictive models. As mentioned in §4.1, this would alter the theoretical foundation of the brain time theory and introduce mechanisms that are not motivated by brain time theoretical considerations. Instead, it would reframe temporal binding as being controlled by top-down expectancies, rather than local processing loops. This undermines the brain time theory's basis.

Moreover, even this non-linear account has the same issues as the linear processing account, in that it cannot account for the many ways in which human perception shows remarkable resilience to temporal distortions, misalignments, and differential processing times. As also noted by White:

> 'An event in one modality can be registered as synchronous with an event in another modality even if they were objectively separated by 200–250 ms (Conrey & Pisoni, 2006; Dixon & Spitz, 1980; Mégevand, Molholm, Nayak, & Foxe, 2013), which suggests a

> substantial processing latency between onset of the first stimulus and occurrence of a synchronized conscious percept.' (White, 2019, 626).

Moreover, cases like the temporal ventriloquism effect (Hartcher-O'Brien & Alais, 2011) show that an auditory signal can shift the perceived timing of a visual event, indicating that perceptual synchrony is not simply a matter of waiting for local re-entrant processing to conclude. Similarly, as noted in §2.1, the study by Holcombe and Cavanagh (2008) shows how prefrontal activity such as selective attentional activity, may retrodictively influence what sensory features are temporally bound. And lastly, as noted in §4.2, causal expectancy (Bechlivanidis et al., 2022; Haggard et al., 2002; Stetson et al., 2006; Stone et al., 2001), repetition (Di Luca et al., 2009; Fujisaki et al., 2004; Vroomen et al., 2004), and exogenous attentional cues (Chennu et al., 2009; Holcombe & Cavanagh, 2008; McDonald et al., 2005), all influence the time represented, without those changes in the time represented being attributable to changes in neural processing times.

These findings suggest that perceptual synchronisation is not solely the result of local re-entrant processing but may also involve a more global, integrative mechanism that dynamically adjusts for processing latencies across different sensory modalities.

The second problem is that although Arstila's explanation may account for the retrodictive effect of the colour-phi motion because it occurs within the 100–150 ms following stimulus onset where re-entrant local perceptual endpoints are determined (Lamme & Roelfsema, 2000), the theory fails to account for other retrodictive effects. There are some cases of retrodiction where it seems that a stimulus presented up to ~500 ms following the onset of some initial stimulus, can affect our perceptual interpretation of that initial stimulus. These long-lasting retrodictive effects present a big problem for the non-linear processing account.

Herzog, Drissi-Daoudi, & Doerig (2020) present a review of such long-lasting retrodictive effects. To highlight one specific example, consider the study by Drissi-Daoudi et al. (2019), in which participants are presented with two sets of stacked vertical lines occurring in sequence, prompting a perception of two diverging streams. When the horizontal lines of one set have a minimally perceptible offset, all other lines in the sequence are perceived as offset in the same direction, even though they are vertically aligned. This offset can be reported by observers, yet it is impossible for them to tell which set of lines is 'carrying' the actual offset. The effect showed a lower bound of approximately 400 ms and an upper bound of 490 ms. This means that a stimulus occurring up to 490 ms after the first stimulus affected perception of the first stimulus. This indicates that motion perception can be disambiguated retrodictively across hundreds of milliseconds (Herzog et al., 2020). A non-linear brain time account would have to claim that the first stimulus ‘bounces’ around the perceptual system for almost half a second before reaching its perceptual endpoint simultaneously with the later stimulus, where these are then integrated. Moreover, because of the thesis of minimal delay, these ~500 ms have to be conceived as the minimal time it takes for this specific stimulus to reach its perceptual endpoint. That is, it cannot be explained away as a way in which the perceptual system artificially delays the perceptual availability of the initial input. This retrodictive effect stretches the account beyond its limits.

A third, more general problem is that the brain time theory provides no explicit explanation of how we often manage to accurately bind features that cooccur in event time as a single temporally unified perception. In many cases, we correctly perceive temporal order and synchrony relations despite neural desynchronisation. Since the brain time theory aims to explain how sensory features are integrated to support temporal perception, it should apply generally to complex multimodal events. This challenge does not fall outside its scope, rather, it is central to its success as a theory of temporal binding.

Arstila (2016a, 180) claims that a consequence of the re-entrant local perceptual endpoints is that they do not necessarily track the events from which they originate. This helps explain the illusion of apparent motion, where the apparent motion has no determinable cause. However, to ensure accurate temporal binding, it seems necessary that the brain has some way of keeping track of the causes and timing of external features. If the relative temporal cues and causal relationships between external sensory features are lost due to neural desynchronisation, then they need to be recovered somehow to allow for accurate event time estimation. This view gives us no explanation of how this is achieved. Such an explanation, however, is necessary if it is to provide a promising general theory of temporal binding that can account for the problem of desynchronisation.

The brain time theory, in general, seems to underestimate the extent to which the brain engages in temporal compensation and recalibration processes when engaged in temporal binding. In the next section, I discuss a final version of the brain time theory that is geared toward modelling how the brain might ensure that multiple features of one event are perceived as a single temporally bound perception despite differential neural processing times.

## 6 Perceptual Endpoints as a Global Integrative Processing Stage

One way to tailor the brain time theory such that it can address the problem of desynchronisation would be by reinterpreting the meaning of perceptual endpoints more radically, as referring to a global rather than a local perceptual endpoint. This would mean that the processing of sensory features cannot be considered finished until some global processing mechanism has processed them in a way that puts them in contact with other processed sensory features from the same event. This postpones the perceptual endpoint to the moment at which all the features of a specific event have finished processing.

The result of such a view would be that sensory features occurring together in event time would reach their perceptual endpoints at the same time, resulting in a temporally bound perception of all these features as being part of one unified perception. This ensures that the brain adapts to differential processing times in order to ensure accurate event time perceptions. This kind of view is even compatible with Arstila's definition of a brain time view, because the 'adaptation [to ensure accurate event time perceptions] is thought to happen before experiences are formed' (Arstila, 2019, 353).

Functionally, this global perceptual endpoint could be modelled in a similar way to the global broadcast mechanism defined by the Global Neuronal Workspace Theory (Dehaene et al., 2006). Under this interpretation, the global perceptual endpoint refers to a global broadcasting activity in which sensory information from specific sensory mechanisms that reaches a certain threshold of neural activation is granted access to other representations in this global workspace (see Mashour et al., 2020). This global broadcasting activity then ensures that content that is broadcast together is represented as occurring together (the time of broadcasting is isomorphic to the time represented).

To ensure that the sensory features originating from the same event are broadcast together, one could hypothesise some kind of time-marker account (White, 2023), according to which signals are time-stamped when they first hit our peripheral nervous system. Signals with the same time stamp then wait for each other to finish whatever processing is necessary, before they reach this global perceptual endpoint together. This would allow for the perception of features that are likely to occur together in event time to also be represented as occurring together in perception.

By definition, this solution also satisfies the thesis of minimal delay. This is because in order for sensory information to become conscious, it must be included in the global broadcast. Thus the minimal delay for a perceptual endpoint is equal to the time it takes to be included in the global broadcast.

This model provides considerable flexibility in ensuring that features from a single event are bound together in a way that maintains their perceived synchronicity. This allows the model to account for a wider variety of retrodictive influences on temporal binding that might otherwise be difficult to explain.

Let me give an example of how this solves the problem of desynchronisation. A touch on the toe will take approximately 200 ms longer to reach the global integration mechanism than a sound (Harrar & Harris, 2005). Consider a scenario in which a sound occurs every 100 ms. Before the sound signal reaches its perceptual endpoint, the global integration mechanism 'knows', by reference to its timestamp, and the fact that a touch on the toe with the same timestamp would take 200 ms longer to finish processing, that it should delay all the incoming sounds approximately 200 ms, in case a touch on the toe, with the same timestamp is incoming.

It is only when all the stimuli that could share timestamps have finished processing that the global integration mechanism delivers its output as a multimodal temporally bound perception. In this case, the global integration mechanism withholds the original sound signal for approximately 200 ms in unconscious processing and uses that signal rather than a newer one that has been processed in the meantime.

With this system in place, brain time theory can ensure accurate temporal binding in ways that compensate for desynchronised neural processing times. I suggest that, if one remains committed to temporal isomorphism and the thesis of minimal delay, one must endorse global perceptual endpoints in order to account for the problem of desynchronisation.

However, as may be obvious, this solution comes at the cost of introducing substantial neural delays between the onset of sensory stimuli and the ensuing perceptual experience. Under this interpretation, perceptions are not made consciously available the moment features are locally processed. Rather, perception lags behind reality by the amount of time it takes for the slowest piece of sensory information to be processed and to play its part in the global broadcast. Moreover, other signals risk being withheld from perceptual availability for no apparent reason.

Of course, it is not the case that we always have to integrate every feature from an event, but even in cases involving only two modalities, waiting for the slowest signal would significantly increase neural delay. This would include delays that make it difficult to account for perception-action synchrony. Moreover, it makes it hard to explain how we use perceptions to guide behaviour in settings where events and features change at fast speeds. Moreover, this model again cannot account for evidence showing that perceptions are sometimes available more quickly than direct sensory processing would allow (see §4.1).

## 7 Moving Beyond the Brain Time Theory

The brain time theory that defines perceptual endpoints in terms of local, feedforward processing fails to account for how we engage in accurate temporal binding despite desynchronisation in neural processing. Two alternative interpretations of perceptual endpoints (Arstila's non-linear processing view and my proposed global perceptual endpoint view) attempt to resolve this issue by redefining perceptual endpoints. While they fare better, the non-linear processing view has difficulties in accounting for the alignment of top-down and bottom-up signals, explaining long-lasting retrodictive effects, and providing an explanation of how accurate temporal binding is generally achieved. While the global perceptual endpoint view may be said to provide a framework that allows it to solve the problem of desynchronisation, this comes at the cost of delaying the perceptual availability of temporally bound perceptions to an unworkable extent, making it rather implausible and unable to deal with the problem of neural delay.

The takeaway from this is that, while the brain time theory presents an empirically interesting and clear theory of temporal binding, it is too simple to address Problem (i) and Problem (ii), set out in Chapter 1. The main issue is that its commitment to temporal isomorphism makes it unable to account for the flexibility in how we temporally bind sensory features under different circumstances. To accommodate this flexibility, one should abandon the brain time theory's thesis of temporal isomorphism. Instead, one should adopt a theory that rejects temporal isomorphism and allows the time of representation to be determined by differential neural processing times without these determining the time represented. This, in turn, allows us to account for how our brain regulates temporal binding across contexts.

Moreover, maintaining the thesis of minimal delay makes it difficult, if not impossible, to account for how our brain compensates for differential neural delays in order to attain more accurate temporal binding, yet such compensations seem to be pervasive when we consider psychophysical effects on temporal binding. As such, there is little hope for developing a satisfactory brain time theory of temporal binding that can address Problems (i) and (ii), while remaining true to its two core theses.

As I will argue in Chapter 3, we can better account for these two problems if we consider temporal binding to be an adaptive and flexible process. In this context, 'adaptive' means that the temporal binding adjusts to differing contextual demands by maximising useful information at minimal costs (Akyürek, 2025). Costs can be understood as the time spent processing sensory information for the purpose of temporal integration, or as useful information lost when perceptual processing prioritises certain features over others. Useful information refers to how temporally integrating sensory features enhances perceptual understanding for some ecological need (Gibson, 1979) including goals, intentions, tasks, judgements, and other behaviours. By ‘flexible’, I mean that temporal binding should allow for variety in what is temporally integrated depending on contextual demands (Akyürek, 2025). Thus, adaptive temporal integration should demonstrate efficiency across varying circumstances.

I suggest that such an account of temporal binding can better be achieved by looking to alternative theories of temporal binding, namely event time theories (presented in Chapter 1). These theories deny both the thesis of temporal isomorphism and the thesis of minimal delay and thus allow the time of representing and the time represented to come apart. In the next chapter, I present an alternative theory of temporal binding that draws inspiration from event time theories but resolves some of the issues that existing event time theories face in accounting for (i) and (ii).

# Chapter 3

## A Hybrid Theory of Temporal Binding

### 1 Introduction

I concluded Chapter 2 with the claim that we must find inspiration in event time theories in order to address two explanatory problems. Problem (i) concerns accounting for the psychophysical data on temporal binding. Problem (ii) concerns accounting for the problem of neural delay and the problem of desynchronisation. According to event time theories, the time of representing (the time at which a perceptual state is formed or realised in the brain) and the time represented (the time in the world that the perceptual state presents as occurring) can come apart. In allowing these to dissociate, event time theories reject the assumption, central to brain time views, that there must be a determinate perceptual endpoint in neural processing that determines the experienced temporal order. In this chapter, I develop an event time theory of temporal binding capable of addressing Problems (i) and (ii).

There are two general strategies for developing an event time theory that breaks the relative ordinality between neural processing times (and thus the time of representing) and the time represented. The first appeals to retrodictive processing. Retrodictive event time theories hold that once sensory signals have been processed to some degree, the perceptual system can retrospectively align how their relative temporal relations are represented. On this view, signals processed at different neural times can nonetheless be experienced as occurring at the same event time because their temporal relations are determined only after they have been processed. Here, the time of representing comes apart from the time represented because the time represented does not directly reflect the time at which the representational content of sensory features has been processed.

The second strategy appeals to predictive processing accounts of time perception (e.g. Blom et al., 2020; Demarchi et al., 2019; Hogendoorn, 2022; Kwon et al., 2015; Nijhawan, 2008; Toren et al., 2020; White, 2018). I describe predictive event time theories as claiming that the perceptual system constructs forward-looking models of how external events are temporally structured and how they are expected to unfold. On this view, temporal binding can occur on the basis of anticipatory representations that are generated prior to the completion of the sensory processing of incoming sensory signals. Whereas retrodictive event time theories allow for the time represented to be determined after the content fixation of sensory features, predictive event time theories allow for the time represented to be determined before the content fixation of the direct feedforward processing of sensory features.

In this chapter, I present both kinds of event time theories in detail and argue that each provides important resources for developing a theory of temporal binding that goes beyond brain time approaches. I show how retrodictive theories can explain how we can retrospectively temporally bind features despite desynchronised or delayed neural processing. I then show how predictive theories can explain how we bind features in ways that allow perception to approximate real-time tracking of unfolding events despite neural delay.

Despite these successes, I claim that both theories face significant issues. Retrodictive event time theories generally fail to account for how we manage to engage in temporal binding in ways that allow us to perceive certain events as occurring in close to real time. By contrast, predictive event time theories only explain how we engage in temporal binding in predictive contexts and fail to account for how our perceptual experiences can be retrodictively updated in unpredictable situations or in situations where later stimuli retrodictively influence earlier perceptual interpretations.

To address these limitations, I propose a hybrid event time theory according to which temporal binding results from an interaction between predictive and retrodictive processes. I articulate this theory via two novel hypotheses.

The first hypothesis holds that predictive and retrodictive processes interact in mutually beneficial ways to bind sensory information in event time. Although hybrid models have been proposed in more restricted contexts, such as trajectory estimation (Grush, 2005; Hogendoorn, 2022), I advance a broader and more general account of temporal binding.

Moreover, even when taken together, predictive and retrospective processing alone does not explain the flexibility and adaptiveness with which we engage in temporal binding. To account for this, I propose a second hypothesis, which claims that attention plays a central role in selecting which contents are integrated into temporally bound conscious percepts. This explains how temporal binding is flexible and adaptive, weighing informational costs against contextual usefulness in light of ecological demands, intentions, and task requirements (for a definition of adaptive and flexible, see Chapter 2, §3).

I argue that this hybrid theory addresses the three explanatory problems introduced in Chapter 1. In this chapter, I focus on how it addresses Problems (i) and (ii). In Chapters 4 and 5, I show how it addresses Problem (iii).

I end the chapter by presenting three big picture consequences about time perception that follow from the hybrid theory. These consequences give rise to further questions, predictions, and explanations of the content and phenomenology of temporal binding that I pursue in Chapters 4 and 5.

In §2, I present empirical evidence that temporal binding is subject to both predictive and retrodictive influences. In §3, I examine retrodictive event time theories (§3.1) and predictive event time theories (§3.2), identifying the limitations of each in accounting for the data presented in §2. This motivates the hybrid theory developed in §4. I introduce Hypothesis 1 in §4.1, Hypothesis 2 in §4.2, and present evidence uniquely supporting the hybrid theory's emphasis on attention in §4.3. In §5, I draw out three major consequences for temporal experience that follow from the hybrid theory presented here. I conclude in §6.

## 2 Psychophysical Evidence for Retrodiction and Prediction in Temporal Binding

To bolster the claim that retrodictive and predictive processing contribute to temporal binding, I provide a range of examples below.

Before doing so, it is important to distinguish predictive and retrodictive *processing* from prediction and retrodiction understood as conscious mental states. We can give a relatively sharp definition of predictive and retrodictive processing in terms of when the brain activates neural representations of the stimulus in question.

Predictive processing can be understood as the pre-activation of neural activity such that the encoding of a neural representation of an expected stimulus precedes the occurrence of that stimulus or at least precedes the feedforward processing of it. Retrodictive processing can be understood as post-activation of neural activity such that the processing of an earlier stimulus is influenced by the processing of a later stimulus, for example, via recurrent processing, feedback loops, or other top-down effects. I use this distinction to categorise the effects discussed below as predictive or retrodictive contributions to temporal binding.

It is far more controversial what it means for a conscious perception itself to be predictive or retrodictive, and there is substantial debate about the extent to which such perceptions exist.

Some predictive processing theorists maintain that all perception is fundamentally predictive (Clark, 2013; Friston, 2010; Hohwy, 2013c; Huang & Rao, 2011), though such accounts have many critics (Colombo & Wright, 2017; Klein, 2018; Litwin & Miłkowski, 2020; Miłkowski & Litwin, 2022). Others have suggested that perception is fundamentally retrodictive (Libet et al., 1979; Libet, 1978), while some remain sceptical that genuinely retrodictive conscious perceptions exist at all, arguing instead that apparent retrodictions can be explained in terms of long integration times due to non-linear processing (Arstila, 2015b, 2016a).

My own view is that conscious perception can be both predictive and retrodictive (that is, the product of both predictive and retrodictive processing), and that in many complex perceptual contexts experience results from an interaction between predictive, retrodictive, and feedforward processing. As such, I will argue that many cases of temporal binding are best understood as hybrid cases in which both kinds of processing play a role.

In what follows, I present what I take to be the strongest empirical evidence that both retrodictive and predictive processing contribute to temporal binding. This evidence helps demonstrate how the ordinality between the neural processing times of sensory features (the time of representing) can come apart from the temporal relations attributed to those features in conscious perceptual experience (the time represented).

### 2.1 Evidence for Retrodiction in Temporal Binding

As described in Chapter 2, retrodictive effects describe visual phenomena where the sensory processing of a stimulus occurring at event time t2 affects the sensory processing of a stimulus occurring at event time t1 and, as a result, affects the conscious perceptual experience of the stimulus occurring at event time t1. Classic examples include cases such as the cutaneous rabbit, where three sets of 3–5 taps on three different locations of the arm are illusorily experienced as ‘travelling’ stepwise up the arm (Geldard & Sherrick, 1972). Another is the colour-phi phenomenon, where two dots flashed in succession at two different locations in two different colours are experienced as a rapid sequence of dots that change colour midway (Kolers & von Grünau, 1976; Ramachandran & Anstis, 1983). Lastly, there are also cases of backwards masking, where a stimulus 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 existence of retrodiction, because they exhibit how subsequent stimuli can affect the perceptual representation of earlier stimuli. Moreover, they have been used to support the idea that the time represented can come apart from the time of representing (see Dennett & Kinsbourne, 1992; Eagleman & Sejnowski, 2000; Grush, 2005; Hogendoorn, 2022; Lee, 2014).

Though, as I said in Chapter 2, §5, the timescale across which the retrodictive integration of stimuli occurs in these cases is often so short (less than 100 ms) that it may be possible to explain these 'retrodictions' in terms of long neural integration times, which would be the result of non-linear processing. Lamme & Roelfsema (2000), for example, argue that the content of perceptual experiences is determined within 100 to 150 ms following stimulation. If this is the case, then it is not clear that the time of representing and the time represented come apart in these cases. Instead, this may simply mean that there is a great deal of subconscious processing (feedforward and recurrent) going on before the sensory features become part of an integrated conscious perception. This means that conscious perceptions that seem to be the result of retrodictive processing, because information presented later influences perception of information presented earlier, might not be cases of truly retrodictive conscious perceptions if they fall below this threshold of 100–150 ms.

However, as I also suggested in Chapter 2, §5, there are cases of much more pronounced retrodictive effects on temporal binding that take place over much longer periods and that cannot be attributed to long integration times or recurrent processing loops.

To get a grasp of how strongly retrodictive processing can influence the temporal binding of sensory signals, consider again the experiment by Drissi-Daoudi et al. (2019) briefly presented in Chapter 2 §5. Drissi-Daoudi et al. (2019) use a sequential metacontrast paradigm where two sets of stacked vertical lines are presented in sequence to prompt a perception 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 perfectly aligned (see Figure 2). The perceived offset can be reported, yet the observer cannot tell which of the set of lines is 'carrying' the actual offset.

The offset 'contaminates' the rest of the lines in the sequence, but it does so in a variety of ways. When attention is allocated to a stream where two offsets in the same direction are present, the perceptual offset is accentuated (Figure 2a). When attention is allocated to a stream where two offsets in opposite directions are present, no offset is perceived, as they cancel each other out (Figure 2b). When attention is allocated to a stream where no offsets are present, no offset is perceived, so the effect is attention-dependent and stream-specific (Figure 2c).

Moreover, these effects of the offset can influence perception up to 450 ms into the past (Drissi-Daoudi et al., 2019). As such it seems that our perceptions seem to remain retrospectively editable for up to at least 450 ms. Herzog et al. (2020) give additional examples of these kinds of long-lasting retrodictive effects.

Sergent et al. (2013) and Thibault et al. (2016) show how post-cueing attention to the location of a stimulus after the offset of the stimulus greatly increases the chance of that stimulus being consciously perceived. Sun et al. (2017) provide evidence of how, even ~300 ms after a sensory input of an ambiguous motion stimulus, people can exert volitional top-down control to affect the perceptual interpretation of the motion direction of that stimulus. McWalter & McDermott (2019) show how we tend to retrospectively ‘fill in’ auditory background information across several seconds when sounds are interrupted by other sounds. Lastly, several experiments by Rimsky-Robert et al. (2019) and Sergent et al. (2011) demonstrate multi-modal retrodictive

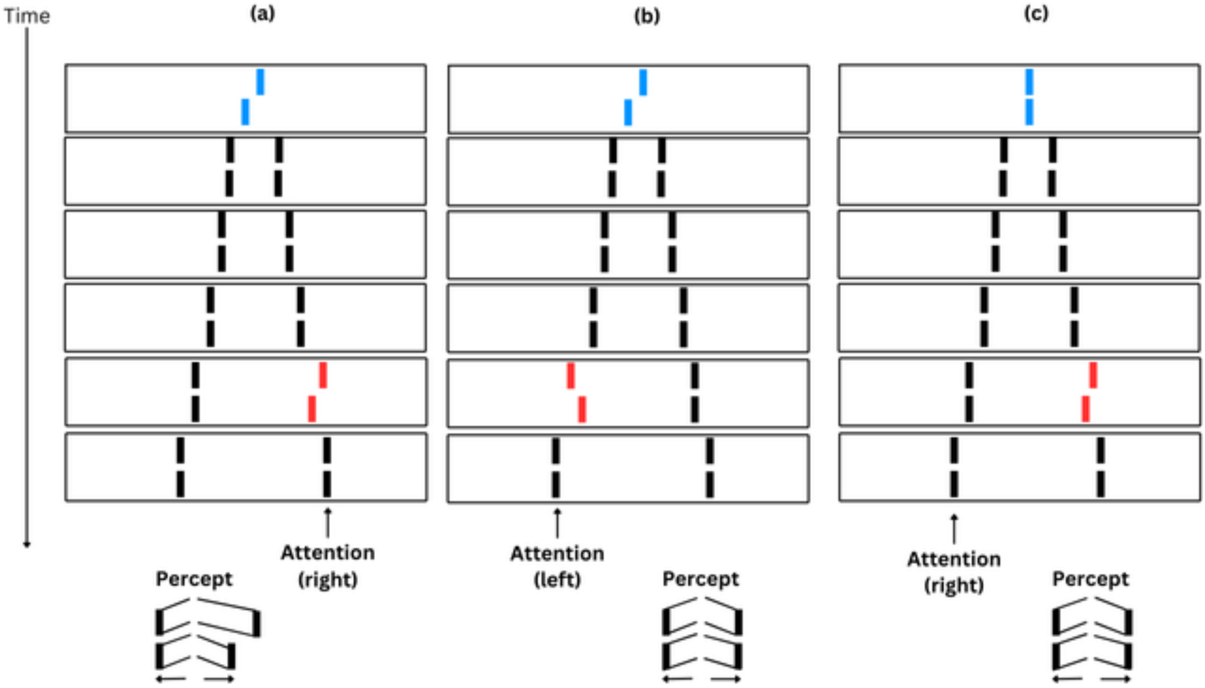

**Figure 2:** Three different ways our perception of streams of stacked lines can be retrospectively affected. This illustration is made to depict the results from Drissi-Daoudi et al., 2019).

effects, showing how near-threshold visual stimuli can be consciously perceived when coupled with a symbolic auditory cue presented up to 200 ms later.

All of these cases support the idea that both bottom-up cues and top-down effects can retrospectively contribute to the conscious perception of stimuli shown much earlier. Clear examples of retrodictive effects are rare because they are difficult to assess experimentally. That is, in many cases they can likely be interpreted not as retrodictions, but simply as perceptions that are the result of long neural integration times (Arstila, 2015b, 2016b). However, these long-lasting retrodictive effects seem to be unequivocally retrodictive and is evidence of how retrodiction plays a role in how we perceptually integrate temporal features over time.

Moreover, just because paradigmatic examples of retrodictive effects on temporal binding (such as those mentioned above) can be interpreted as not being genuine conscious retrodictive perceptions, does not mean that they are not retrodictive perceptions. This is simply an area where we need more rigorous psychophysical research that tries to tease apart retrodiction from long integration times.

As I discuss below (§3.1), appeals to retrospective temporal processing help explain how our perceptual system compensates for differential neural latencies. Specifically, retrodictive processing can contribute to temporal binding in the sense that, when we are faced with an event composed 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 perceptual event.

### 2.2 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; Dołęga & Dewhurst, 2020; Friston, 2010; Hohwy, 2013; Knill & Pouget, 2004; Parr & Friston, 2017; Summerfield & de Lange, 2014). In this section, I present some of the evidence in support of how prediction may contribute to temporal binding. Explicitly spelling out this relation between temporal binding and prediction is becoming increasingly important, as prediction has significant implications for time perception.

In the case of temporal binding, prediction can be used to accommodate discrepancies between processing times of various sensory features and accommodate the delays involved in neural processing. That is, if external sensory features' temporal relations can be predictively integrated with each other, then there is no need to retrodictively align them after they have all been processed. Moreover, there is no need to delay perceptual availability by the time that such processing lags behind reality, meaning that our temporally bound perceptions can be consciously experienced well in advance of the otherwise delayed direct processing of sensory features.

To motivate the various roles of predictive processing, I list and discuss evidence for four different ways in which predictive processing influences our perception of how sensory features are temporally bound. These include (i) motion extrapolation, (ii) predictive updating, (iii) predictive overpowering, and (iv) mnemonic expectations.

(i) ***Motion extrapolation****:* One example that is often considered a direct result of feedforward prediction is motion extrapolation, first proposed by (Finke et al., 1986), which involves using information about some features' or objects' past locations (trajectory) to infer their current and next locations. As such, it is also often referred to as trajectory estimation. There are several predictive accounts of motion extrapolation (Blom et al., 2020; Johnson et al., 2023; Kimura, 2021; Nijhawan, 1994, 2008; Turner et al., 2024, 2025) and they are often used to account for a series of motion illusions where people tend to perceptually overshoot the location of a moving stimulus, judging it to be further along its most likely trajectory than it currently is. This includes motion extrapolation accounts of the flash-lag effect (Hogendoorn, 2020; Holcombe & Corbett, 2023; Nijhawan, 2008; Rust & Palmer, 2021; Teichmann et al., 2021)[9] as well as of representational momentum (Kimura, 2021; Ruppel et al., 2009).

The flash-lag effect is a visual illusion where a briefly flashed object appears to lag a moving object, even when they are aligned. The central idea of the predictive motion extrapolation account is that this perceptual misalignment arises because the brain processes the next expected position of the moving object. This kind of extrapolation cannot take place for the sudden flash, so we see the moving object as ahead of the flash. In a recent review by Hogendoorn (2020), he examines 25 years of research on this effect since Nijhawan's (1994) predictive motion extrapolation explanation and argues that 'the accumulated evidence from neurophysiological, computational, behavioural, and neuroimaging studies converges to support the existence of motion extrapolation mechanisms in the visual system, as well as their role in generating the FLE [Flash Lag Effect]' (Hogendoorn, 2020, 5703).

---

[9] For more examples see Holcombe & Corbett (2023) who have argued that while predictive motion extrapolation likely contributes to the flash-lag effect it is unlikely that there is any such single-cause explanation of the effect. They present a series of counterexamples to the predictive motion extrapolation account and present a range of temporal error mechanisms that may also contribute to effects like the flash-lag effect (I discuss this more in §4.2.4).

Representational momentum occurs when observers viewing a moving or rotating sequence tend to remember the final position shown as slightly farther along its path than it is. In reporting the last seen angle of a rotating object people often choose a probe that reflects the 'next' implied frame rather than the actual one (Hubbard, 2014). Kimura (2021) argues that representational momentum should be explained as a case of predictive processing. It is by predictively modelling the momentum and trusting this prediction that participants fail to choose the correct probe. Several indicators suggest 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, supporting the idea that representational momentum is an automatic perceptual consequence of ingrained predictive inferences.

Taken together, these studies suggest how extrapolation occurs during early visual (retinal) processing, while other studies (Blom et al., 2020, 2021; Johnson et al., 2023; Turner et al., 2024) show that extrapolation occurs across later (cortical) visual stages. All these kinds of extrapolation cause visual perceptual representation to shift closer to event time. As such, motion extrapolation is a clear case of how predictive processing can contribute to temporal binding of features across different levels of processing, thus overcoming delays that would otherwise occur if our perception were based on direct processing with no extrapolation.

(ii) ***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 who are repeatedly presented with a sound and a flash 235 ms apart will, within 3 minutes of exposure, shift their point of subjective simultaneity[10] by approximately 22 ms to compensate for the offset, thus making them more likely to see the flash and sound as simultaneous when the flash and sound are moved closer together in time (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).

---

[10] The temporal offset between two stimuli at which an observer perceives them as occurring simultaneously

According to Holcombe (2015, 830), the reason why our perceptual system shifts 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 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.

(iii) ***Predictive overpowering****:* In a recent study Bechlivanidis et al. (2022) show how our temporal 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 C-B to the perceived sequence B-C. Participants are fully attending and allowed to view the sequence repeatedly, yet the reversal effect persists. This reversal effect indicates that there are predictions that overpower the external visual signals. These results are easily interpretable if we take perception to make use of 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 causal inference based on complex ingrained memories of causal structures that influences the perceptual extrapolation effect. The extrapolation is not due to the effect of short-term memory or some feedforward processing of the stimuli. 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. In the case of 'temporal ventriloquism' we bind information relative to salient gestalts (Hartcher-O'Brien & Alais, 2011; Vroomen & De Gelder, 2004), in other situations our perceptual system compensates for the relative differences in the propagation speed of sound and light (see Alais & Carlile, 2005), and in some cases, we also see events as closer together if we take ourselves to be the cause of the event (Haggard et al., 2002; Stetson et al., 2006). In all these situations, some top-down causal representations are necessary for the effects to take place (see Hoerl et al., 2020, 23; Tramacere & Allen, 2022, 14).

(iv) ***Mnemonic expectations***: When trained to learn a sequence of images, actions, or associations between two stimuli/actions, people become very proficient in perceptually preparing for expected stimuli, resulting in much faster perception of the stimuli in question or the action to perform (Hindy et al., 2016; Monno et al., 2022; Tamir & Thornton, 2018; Thomaschke et al., 2015; Thornton & Tamir, 2020; Volberg & Thomaschke, 2017). Such expectations not only activate the same areas as feedforward visual predictive mechanisms but also feedback activity, such as the retrieval of past experiences. Hindy et al. (2016) 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, i.e., activates some ingrained pattern of inferential processing. This shows how learned associations and expectancies can facilitate temporal binding of complex events based on higher-order feedback mechanisms.

With these examples, we see how predictive processing influences temporal binding by displacing direct processing with predictive processing of temporal relations. These predictions affect both the time of representing and the time represented. Predictive processing can shift the time of representing closer to event time. This helps the perceptual system to represent features in due time despite neural delays in direct processing. Moreover, predictive processing can be used to predictively model the time represented by modelling the expected temporal relations of sensory features.

## 3 Retrodictive and Predictive Event Time Theories

Though the evidence that our time perception is influenced by retrodictions and predictions is growing, there is little consensus about how we should go about providing a coherent explanation of both within a theory of temporal binding. Below, in §3.1, I introduce two event time frameworks by Dennett & Kinsbourne (1992) and Nishida & Johnston (2010) that explain temporal binding via retrodictive processing. In §3.2, I propose how one might formulate a predictive (rather than retrodictive) event time theory of temporal binding using the predictive processing framework proposed by Clark (2013). I propose some benefits and shortcomings of both these kinds of views and argue that we can get the best of both worlds with the hybrid theory proposed in §4.

### 3.1 Retrodicting Time

Retrodictive event time theories take it that the order in which features are processed in specific mechanisms (content fixation) can be retrodictively broken so that the content of our temporally bound perceptions better represents the event time of external sensory features.

This can be accomplished by having a perceptual system with 'time markers' that timestamp sensory signals' event time by encoding signals' sensory onset time (Nishida & Johnston, 2010). Any desynchronisation that might take place as these signals are processed in different mechanisms can then be compensated for by retrodictively aligning the event time of processed signals with reference to the encoded time markers. This kind of 'time marker account' has been set out by Libet et al. (1979) and further developed by Nishida & Johnston (2001, 2010). A distinct theory by Dennett and Kinsbourne (1992) posits a system where any desynchronisation in the processing of sensory features is compensated for through temporal binding being sensitive to contents. According to them, temporal binding is done with regard to how well encoded representational contents temporally fit together relative to a range of top-down goals and bottom-up contexts. This is the 'content-sensitive settling view'.

### *3.1.1 Nishida & Johnston time marker view*

As also mentioned in Chapter 2, Nishida and Johnston's time marker account (2010) relies on a series of mechanistic evidence about feature processing to claim that the brain groups the timing of various sensory signals by having two kinds of mechanisms in place: early-level mechanisms and mid-level mechanisms.

The early-level mechanisms mark the onset of sensory signals by tracking the transients of first-order changes (colour change, position change, luminance change). These are quick, feature-specific mechanisms with high temporal resolution. Because these early-level mechanisms work quickly and pick up on the onset of intramodal sensory signals, they represent something close to event time.

These early-level time markers are then used as input to mid-level mechanisms, which operate more slowly and in an amodal, cross-modal manner. The mid-level mechanisms are responsible for integrating features over time, particularly representing second-order changes, such as motion reversals or other continuously changing stimuli that do not produce unique transients. As mentioned in Chapter 2, the mid-level mechanisms must compare at least three successive points in time to determine the location of second-order changes. It may, for example, involve comparing an initial location of a moving stimulus to a later location to determine direction and then comparing to a third location to determine whether the trajectory continues or changes. Mid-level mechanisms achieve temporal binding by selecting, individuating, and registering the time markers of salient first-order features and performing a 'saliency-based cross-channel comparison' (Nishida & Johnston, 2010, 284), thereby associating second-order changes with early-level time markers.

Temporal binding that combines first-order and second-order changes is inherently imprecise because these changes require different processing times and rely on different mechanisms, making direct temporal comparison between them challenging.

Nishida & Johnston's account allows parallel sensory mechanisms to operate in desynchronised ways while compensating for differential processing delays. Early-level mechanisms provide precise event times for salient transients, while mid-level mechanisms integrate other signals over time by aligning features across modalities based on these time markers.

While Nishida & Johnston's account may be applied more widely, the proposed mechanisms are developed as a way to provide a better theoretical explanation of the colour-motion asynchrony illusion compared to existing brain time accounts that explain this in terms of synchrony in neural processing times. I presented these competing accounts of this illusion in §2.1 (Chapter 2). For the purpose of this chapter, the important takeaway from this account is that it allows for a parallel system of sensory mechanisms to fulfil their perceptual tasks in desynchronised ways while compensating for any differential processing delays that take place in the further processing of these sensory features by temporally binding features according to their event time.

### *3.1.2 Dennett & Kinsbourne's Content-sensitive Settling View*

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

The multiple-drafts model is a theory about how contents become cognitively accessed, i.e., available to guide reasoning, reports, deliberation, and consciously guided actions. In this view, our nervous system encodes an abundance of content 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 accessed).

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 provides a functional account of how contents gain conscious control. As its name indicates, this selection process is sensitive to representational contents, rather than to the timing of the vehicles that represent them. This means that the competition value of contents is determined by how salient they are and how well they fit relative to top-down perceptual goals, expectations, intentions, task demands, and bottom-up context.

According to this view, any conscious perception 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. A conscious perceptual experience can be re-drafted over time without us being aware that this redrafting has taken place. The re-drafted perception may completely overwrite earlier representations, to the extent that there is no way to tell whether there has ever been a first draft once it has been overwritten.

Dennett and Kinsbourne (1992, 191) propose two distinct ways in which one could think that such redrafting takes place. They call these Orwellian and Stalinesque revisions. Consider the classic example of retrodiction called the cutaneous rabbit (Geldard & Sherrick, 1972). In this experiment, a person is tapped 3–5 times on three different locations on their lower arm. When asked about their experience, participants usually report that it felt like the taps were 'travelling' stepwise up the arm, rather than being experienced in three distinct locations.

According to the Orwellian interpretation we initially experience the first set of taps as being located at the same locations, but then when we experience the next set of taps at a location further up the arm, our memory of the initial veridical experience is rewritten so that we instead remember it as if the taps were travelling up the arm. So it is only our memory and not our experience that is redrafted, but this is all done before we can report anything about the experience, which is why we report that we experienced the taps as travelling up the arm.

According to the Stalinesque interpretation our experience of the first set of taps at the first location is rewritten in light of the second set of taps at the second location before we become conscious of the experience of the initial taps. Thus, we have an illusory experience of the taps as travelling up the arm. On this account, the content of our experience is rewritten before we get to consciously experience it.

Dennett & Kinsbourne (1992) propose this distinction only to deter people from making it. That is they think that it is nonsense to distinguish between the Orwellian interpretation and the Stalinesque interpretation in the case of temporal binding. This is because drawing such a distinction requires that there is a determinate moment at which a perceptual experience becomes conscious. However, if our experiences are continuously subject to revisions in this way that depend on which of multiple drafts is currently in the limelight, it seems that there will always be multiple possible answers to what the contents of our experiences are at any specific moment. Moreover, our only way to get access to the content of someone's perceptual experience is to elicit some kind of behavioural 'report' from the person in question, but whatever is reported will underdetermine the choice between the Orwellian and the Stalinesque interpretation.

I tend to agree partly with this analysis, in particular with the claim that the choice between the two interpretations is, as of yet, underdetermined by the available data (even now, 34 years on). However, I think that the distinction is sensible in that it may motivate interesting empirical studies that intervene on processing streams at different times, relative to assumptions about how the conscious perception of sensory features develops, in order to examine how such interventions influence perceptual interpretation. When this is said, matters are perhaps empirically even more messy than Dennett and Kinsbourne suggest. This is because, as I claim in §4, predictions and not only retrodictions influence our perceptual interpretations of what happens when; perhaps both things are happening. In a seemingly Stalinesque manner, information from different modalities seems to perceptually integrate and interfere with each other in early cortical processing (Groen et al., 2013), expectations influence our predictive extrapolation of locations both during retinal and cortical processing (Blom et al., 2020, 2021; Johnson et al., 2023; Turner et al., 2024). Yet in an otherwise seemingly Orwellian fashion, our perceptual processing (and perhaps conscious perception) of an initial stimulus, can be retrodictively reinterpreted up to 490 ms after its onset (Herzog et al., 2020). I have no particular stake in whether one, both, or neither of these interpretations are always or contingently true, but I do think that Dennett & Kinsbourne are right that it is temporally indeterminate exactly when a temporally bound experience is formed. I return to this in §5.1.

For now, what is important is that the contents that are temporally bound into a conscious perceptual experience are those that together provide the best representation of the environment according to current top-down goals and bottom-up context and inputs (Dennett & Kinsbourne, 1992, 197). Dennett and Kinsbourne (1992, 189) describe this view as 'a cheaper, less foolproof but biologically more plausible' alternative to the kind of time marker view implied by Nishida & Johnston (2010) (though Dennett and Kinsbourne's target is Libet's (1989) time marker model). This is because, rather than tracking explicit time markers for each content, the system can infer a 'relative phase of best correspondence' among available contents. In this way, contents can be accurately temporally bound without requiring explicit time markers[11]. Moreover, on Dennett & Kinsbourne's model, unlike Nishida & Johnston's model, our brain does not need to represent the temporal relations of sensory contents twice: first as time markers at sensory onset and then again when they are consciously experienced. On Dennett & Kinsbourne's model, our brain only represents temporal relations when we consciously experience them.

Despite this, Dennett and Kinsbourne (1992) are very vague about the specific mechanisms underlying content-sensitive settling (Akins, 1996; Mölder, 2014). The hybrid theory presented later can be seen as an extension of their model that is intended to replace some of this vagueness with more detail.

A key difference between the content-sensitive settling account and Nishida and Johnston's time marker account is that Dennett & Kinsbourne view temporal binding as more flexibly influenced by top-down goals and bottom-up context. Temporal binding is not necessarily determined by encoded temporal markers, as in Nishida & Johnston's view.

This flexibility allows their model to account for a wider variety of retrodictive influences on temporal binding that may be difficult to explain using Nishida & Johnston's framework. For example, it is unclear how the time marker theory could account for the retrodictive integration of line segment offsets across hundreds of milliseconds (see Figure 2).

However, the explanatory breadth of Dennett and Kinsbourne's theory comes at the cost of explanatory depth. According to Nishida and Johnston (2010, 295), their account is too simple. For example, '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 systematically 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 experimental settings, and the model provides no distinct mechanisms for generating detailed predictions[12].

Considering the different scopes of the theories, it may be useful to think of Nishida & Johnston's view as a narrow neurophysiological implementation of Dennett & Kinsbourne's

---

[11] 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.

[12] Conceptually similar models like that of Grush's (2005) or Hogendoorn (2022) might fare better because they actually provide much more distinct predictions.

broader functional account. In tracking the timing of first-order transients, the brain may employ time markers, or it may not. The important functional agreement between the theories is that the contents of the time represented are not strictly determined by the ordinality of neural processing, nor is the time represented settled millisecond by millisecond. Rather, time is settled by a combination of the event time of features, the saliency of those features, and the retrospective temporal alignment of their representations.

Nishida and Johnston may be correct in positing time markers, and Dennett & Kinsbourne may be correct that the represented time is not fully determined by such markers. Both theories remain open on these points. While Nishida & Johnston (2010, 296) explicitly describe early-level mechanisms tracking the onset of luminance transients, they are vague about how mid-level mechanisms extract salient features and compare them across modalities. They say that this extraction selects for salient changes, something which is likely influenced by top-down effects, and which could affect what features are temporally bound in different circumstances. Similarly, Dennett & Kinsbourne (1992, 189) remain open to the use of time markers but prefer a system without them, avoiding the double representation of sensory timing at onset and at conscious experience, which they argue are biologically superfluous.

Thus, the main insight of both theories is compatible: the perceptual system is retrodictive and influenced by top-down and bottom-up saliency cues. They differ primarily in scope and explanatory goals. Dennett & Kinsbourne aim to account for temporal binding more generally, including natural settings, whereas Nishida & Johnston (2010) focus on providing a mechanistic account of the colour-motion asynchrony illusion.

### *3.1.3 A Neural Delay Issue for Retrodictive Event Time Theories?*

When it comes to accounting for the problem of desynchronisation and the problem of neural delay, both theories perform well in addressing desynchronisation. However, it may seem that retrodictive theories face a challenge in explaining how the brain compensates for neural delays. According to these theories, the time of representing is delayed by the time it takes the brain to process all relevant signals of an event and retrospectively align them. This would presumably introduce significant delays in our perceptual experience.

As such, retrodictive event time theories cannot fully explain perceptual experiences in which temporally bound features are experienced as occurring in close to real time. Consider, for example, the case of trajectory estimation from §2.2. In this case, extrapolation occurs during both early visual (retinal) processing and later cortical processing, enabling us to perceive visual trajectories in close to real time and thus overcome delays that would otherwise occur if perception were based solely on direct processing (Blom et al., 2020; Johnson et al., 2023). Retrodictive event time theories cannot account for this, because information about the occurrence of sensory features still lags behind the world by the time it takes for features to reach the peripheral nervous system and be processed by early low-level sensory mechanisms.

I argue, however, that this issue can be addressed by integrating retrodictive event time theories with what I call predictive event time theories. I will turn to describing this kind of theory next.

### 3.2 Predicting Time

Predictive processing has been proposed as a grand unifying theoretical framework that promises to account for all mental phenomena and cognitive functions (Clark, 2013; Hohwy, 2013; Ransom et al., 2020) including *action* (Clark, 2016, 2018; Friston et al., 2007; Friston et al., 2010; Hohwy, 2013b, 2013a), *perception* (Andy Clark, 2013; Hohwy, 2012; Hohwy et al., 2008; Kiebel et al., 2009; Sprevak & Smith, 2023), *consciousness* (Clark, 2019; Clark et al., 2019; Hobson & Friston, 2014; Hohwy, 2012), *emotion* (Hesp et al., 2021; Ransom et al., 2020; Van de Cruys, 2017), and *sensory processing* more generally.

While there are plenty of concerns that predictive processing theories may not deliver on these large-scale promises (Colombo & Wright, 2017; Klein, 2018; Litwin & Miłkowski, 2020; Miłkowski & Litwin, 2022), there is little doubt that predictions at least contribute to the integration of temporal information, and that this is achieved through some form of predictive processing (see §2.2). Below, I provide a brief overview of the predictive processing framework in order to show how such an account can be used to formulate what we can call a predictive event time theory (an event time theory with predictive processing as its underlying mechanism). This is meant as an instructive overview of how predictive processing may contribute to temporal binding more generally.

At a general level, predictive processing theories aim to solve how the brain connects to its external environment when it can only ever be aware of its own internal states and processes, and how these change over time. 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 its own responses to these signals and the subsequent effects thereof. As such, the brain does not passively register incoming signals, rather it actively predicts them within a hierarchical system of processing.

By passing informational messages up and down different layers in the hierarchy, the brain can represent an overview of the matrix of hidden causes underlying the sensory data transduced by various systems in the hierarchy and ultimately by our peripheral sensory receptors (Badcock et al., 2019; Friston, 2008; Lee & Mumford, 2003). Each layer in the hierarchy attempts to predictively model the activity of the layer below while being predictively modelled by the layer above. Errors in predictions are sent up the hierarchy, and the goal of each layer is to minimise the number of prediction errors sent upward. Input signals in early-level sensory mechanisms lower in the hierarchy thus fuel the generation of representational models at higher levels in a bottom-up way. These representational models aim to accurately predict incoming signals and, in turn, can inform early-level perceptual processing in a top-down fashion.

A function of this hierarchical system is that higher levels learn to track the structural activity, rather than just the actual activity, of the level below at specific temporal regularities. This means that higher levels tend to track slower-changing and more abstract regularities compared to lower levels (Harrison et al., 2011; Kiebel et al., 2009).

This predictive system can be characterised as generative because it relies on encoded expectations about how hidden external causes give rise to sensory inputs to generate predictions, which are continually checked against incoming sensory information (Friston, 2010; Sprevak & Smith, 2023). This allows the perceptual system to track the state of the world by continuously attempting to replace incoming sensory signals with encoded predictive models.

A final important aspect of the predictive processing framework is the way in which the perceptual system can actively minimise prediction errors by either changing its own internal states (modifying perceptual models) or the sensory inputs it receives through acting on the world (changing actions) (Adams et al., 2013; Kaplan & Friston, 2018; Parr, Corcoran, et al., 2019). For example, we can act in certain ways (including moving our bodies or shifting our attention, consciously or unconsciously) to sculpt the world, or at least the stimulus array, to fit our predictive models and minimise prediction errors. Alternatively, we can adjust our perceptual models to fit how we expect the world to change as a result of certain actions. In this sense, the brain not only predicts environmental changes but also predicts how our active engagement with the environment will affect the sensory input and how we can act to minimise prediction errors.

### *3.2.1 A Predictive Event Time Theory*

With the predictive processing framework in view, we can now outline what I call the predictive event time theory. This model of temporal binding is grounded in hierarchical predictive processing. On this account, temporal binding is not merely a consequence of passively registering the order in which sensory signals arrive. Rather, it is the outcome of the brain's use of predictive generative modelling of the temporal relations of sensory features to integrate temporal information about when they are expected to occur relative to each other, ultimately integrating them into temporally bound perceptions. This system relies on prior statistical information to establish reliable inferences, which can be used for top-down modelling of the expected temporal relations of external sensory features, while sending information about errors in these predictions back up the hierarchy in a bottom-up fashion.

Since this predictive system aims to replace direct processing with predictively processed perceptual representations, a well-functioning system would allow us to temporally bind sensory features before the sensory signals would have otherwise finished processing had the system relied solely on direct feedforward processing. In other words, when the perceptual system has a strong predictive grasp on the temporal relations of external sensory features, it enables both accurate and speedy temporal binding. Engaging in temporal binding in this way provides a clear advantage over a system relying only on direct processing and is consistent with the evidence presented in §2.2.

As these studies indicate, there are a lot of situations where it seems that predictability influences our perceptual interpretation of which features are temporally bound. In fact, there is already substantial literature providing predictive processing accounts of temporal phenomena such as causal relations (Clark, 2013), social actions (Bach & Schenke, 2017; Kilner et al., 2007; Thornton & Tamir, 2020), trajectory estimations (Blom et al., 2020), and representational momentum (Kimura, 2021). Such perceptions are otherwise difficult to explain through direct processing of external stimuli due to factors such as neural delays, desynchronised processing, impoverished stimuli, or ambiguity in the sensory input.

The main theoretical limitation of predictive temporally bound perceptions is that they are only useful if they are accurate. Speedy predictions about temporal binding are maladaptive if they are misleading. If the perceptual system produces many such inaccuracies, it should lead us to distrust our own perceptions. Therefore, inaccurate predictive 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 predictive event time theories.

## 4 A Hybrid Theory of Temporal Binding

With this rough sketch of the retrodictive and predictive effects and how they can be used to develop distinct event time theories of temporal binding, I will present and argue for two hypotheses about how predictive processing can be integrated with retrodictive processing views. The first concerns how retrodictions and predictions can interact in mutually beneficial ways to ensure speedy and accurate temporal binding. The second concerns a description of how attentional modulation plays a central role in the flexibility and adaptability with which contents are temporally bound.

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

‘The reason for the vigorous debate between prediction and postdiction [what I call retrodiction] 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, 2008a, 205).

I support this sentiment, and while Eagleman accurately emphasises why we need a hybrid theory, he leaves it completely open whether there could be any theory that integrates these two sides of the story and what such a theory would look like. By contrast, Grush (2005) and Hogendoorn (2022) do provide substantial hybrid theories that fit well with the hybrid theory I propose. However, there are several differences to note between my theory and their theories.

First, 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 hypothesis I propose is broader than the trajectory estimation views of Grush (2005) and Hogendoorn (2022).

Second, I do not agree with parts of Grush’s (2005) and Hogendoorn’s (2022) theories, which imply that we perceptually represent whole timelines and trajectories. Sometimes we may simply experience content representing an isolated moment. Despite being subject to predictive drafting and retrodictive redrafting, this does not constitute a timeline representation. Moreover, to the extent that we may represent a certain timeline of how sensory features are temporally ordered, we only do so in a very limited sense, by representing diachronic temporal relations of some subset of sensory features that we attend to over time. These timelines would also change continuously as the environmental context or our top-down attention changes. I discuss this point at length in Chapter 4.

Third, my second hypothesis makes a substantial novel claim about the role attention plays in facilitating this hybrid interaction between predictive and retrodictive temporal binding and in determining which contents we become conscious of and which contents remain unconscious. It is this component that allows the theory to explain the high degree of flexibility and adaptability that our temporal binding capacities exhibit. It is also from this claim that a series of consequences about temporal experience and empirical predictions about temporal binding follow, which I lay out in §5 and in Chapters 4 and 5.

### 4.1 Hypothesis 1: Predictive drafting and retrodictive redrafting

By integrating retrodictive event time theories (§3.1) with the predictive event time theory outlined above (§3.2), we can address the issues set out for each theory.

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 with 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, 2025) show how such extrapolation activity can be found during both early visual (retinal) processing and a series of later stages of cortical visual processing. Using EEG to track neural activity, they found that when people are tracking the visual trajectory of an object, the brain's processing of the stimulus' locations is shifted forward across temporally distinct processing stages, 'with representations progressively shifted closer to real-time' (Turner et al., 2025).

Other evidence of the existence of relevant predictive mechanisms comes from EEG and fMRI decoding studies of predictive 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). All these extrapolation mechanisms cause the representation of time to shift closer to the occurrent event time of the features in question.

By integrating such extrapolation mechanisms within the structure of existing retrospective event time theories, they can account for various predictive effects on temporal binding. Take the ‘content-sensitive settling view’. Within this model, we can take the temporal contents that are settled into conscious drafts to be partially driven by predictive models of the ‘relative phase of best correspondence’, based on ingrained hierarchical inference patterns, rather than being the result of a purely retrodictive process. In this case, the ‘relative phase of best correspondence’ would be partially driven by the best predictive model of the temporal relations of external events. The temporal binding of sensory features can be driven by predictions of sensory features’ temporal relations, in turn making them less stimulus-driven and less affected by differential neural latencies of such direct processing. As such, integrating these mechanisms contributes to mitigating the problem of neural delay that retrodictive event time theories cannot easily account for (see §3.1.3). On this hybrid version, our system does not merely continuously redraft conscious perceptual interpretations as more information is processed but may also redraft conscious perceptions ahead of such processing and draft the phase of best correspondence based on predictive models.

What about the problem with the predictive event time theory (§3.2) that the usefulness of predictions is constrained by their accuracy, meaning that they 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 unexpected stimuli. It would be dire if our system constructed temporal perceptual predictions that turned out to be very inaccurate, especially if we could have avoided this simply by waiting the milliseconds it would take to temporally integrate sensory features based on direct feedforward processing. Moreover, if such predictive errors are common, they would likely create a sense of distrust in our own perceptions of temporal relations of order and simultaneity as we cannot reliably tell which are predictive and which are not.

By integrating a predictive event time theory with retrodictive event time theories we 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 §3.1.

If prediction errors are quickly corrected in a retrodictively revised conscious draft (see §3.1.1 for a definition), then there is no major issue with prediction errors. We might even just forget them, or never fully experience the erroneous predictive perceptions in any way that allows us to consciously act on them (see Stalinesque/Orwellian distinction §3.1.1). 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. EEG activity in trajectory estimation tasks shows evidence of ‘representational overshoot during early-stage processing following unexpected changes to an object's trajectory', and that this same trend occurs in unsupervised simulated neural networks (Turner et al., 2025)

While predictive processing allows for shifting representation closer to real time, this predictive trajectory perception should ideally be ignored 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 perceptually overshoot the turning point of an unexpectedly reversing trajectory. It might be that this is what happens. Maybe we briefly perceive the object ahead of the location where it reverses and then update our predictive model accordingly in light of incoming prediction errors. This would, of course, involve a lag in processing the transients following reversal and updating our predictive model accordingly, and it would likely involve a perception of the object as either reversing from the overshot location with extra speed to catch up (smoothing over the prediction error), or a brief jump in the trajectory to catch up to the updated prediction.

However, this does not match what we perceive. In these cases, observers generally do not report perceiving the moving object beyond the location where it reverses unexpectedly (Eagleman & Sejnowski, 2000). In manipulated versions of representational momentum or the flash-lag effect, where the moving object reverses at the time of report instead of continuing, people report a locational displacement in the new reversed direction of motion (Chappell & Hinchy, 2014; Hubbard, 2005)[13].

Other studies also indicate that when the trajectory continues in the unexpected direction, we seem to perceive sudden motion reversals in close to real time, despite the fact that EEG data indicate that we briefly pre-activate neural encoding of the object's expected location on the initial trajectory (Blom et al., 2021). Blom et al. (2021), however, also show that while the initial latency advantages gained by predicting the trajectory were lost in the case of an unexpected motion reversal, 'the latency advantages for subsequent sequence positions were restored almost instantaneously' (Blom et al., 2021, 200). The motion reversal predictions are updated to reinstate the latency advantage, and they suggest that this is seemingly done in a retrodictively revisionary way (e.g., by signalling motion reversals, see Schwartz et al., 2007) to ensure perceptual smoothing of the motion trajectory of the stimuli. Some kind of retrodictive mechanism thus seems to be in play in the case of motion extrapolation. In this case, retrodiction might simply mean that the temporal window over which the trajectory is perceptually interpreted is extended, to ensure that

---

[13] This seems to depend partly 'on the observers' prior expectations about the object and its trajectory (Verfaillie and d'Ydewalle, 1991; Reed and Vinson, 1996)' (Hogendoorn 2020, 5703).

our conscious perception of the initially predicted location of the non-reversed trajectory is overwritten by a redrafted interpretation.

This kind of retrodiction is similarly needed to account for the long retrodictive effects presented in §2.1. To account for the fact that our perceptual experience of a stimulus occurring at time t1 can be retrodictively affected up to 450 ms 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 predictive perceptions are simply updated in light of incoming prediction errors, we should only experience the offset of the lines after the offset stimulus that prompted a prediction error. But again, this is not what happens. Our perceptual experience before any stimulus that could prompt a prediction error is also ‘contaminated’. A similar story might be told of other cases presented in §2.1 where our conscious perceptions rely on retrospective processing due to the stimuli being rather unpredictable, hence requiring further processing.

By integrating retrodictive redrafting 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 have 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 accurately represented yet delayed time of representing in the case of stimuli that violate predictions. As such, perception based on prediction errors should be fleeting and forgettable if new information or context engages retrodictive redrafting. Such perceptions might simply 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.

While this kind of integration of retrodictive and predictive event time theories might seem to get us a long way in providing a functional account of temporal binding, it does not say anything particular about how these kinds of processing interact. That is, how does our perceptual system ‘know’ to rely on predictively encoded information, or to engage in continuous retrodictive reinterpretation of our perceptual conclusions about the temporal relations of sensory features? Moreover, the variability across psychophysical studies in how we engage in temporal binding seems to indicate that there is a great deal of flexibility and adaptability involved in what features we perceive as temporally bound, and when. This kind of interaction and variability needs to be explained. In the next section, I argue that attentional modulation of temporal contents plays a key role.

### 4.2 Hypothesis 2: Attentional Selection of Content for Temporal Binding

The second hypothesis of the hybrid theory accounts for why some contents get temporally bound into conscious perceptions, while other contents remain unconscious and unbound. 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 which contents get temporally bound.

I take this appeal to attention to be motivated by a range of different results that indicate how attention plays a central role in temporal binding. See points (i)-(iv):

(i) Predictions that contribute to visual tracking of moving objects are capacity-limited, suggesting that these predictions are attention-dependent (the predictive edge only applies to two objects at a time) (Howe & Holcombe, 2012). Moreover, there is preliminary evidence that predictive motion extrapolation is attention-dependent (Hogendoorn, ACNS 2025) and only takes place when it is needed for some judgments or action (Gastrell, ACNS 2025).

(ii) Almost all evidence of retrodictive effects on perception relies on bottom-up or top-down modulation of attention (Drissi-Daoudi et al., 2021; Rimsky-Robert et al., 2019; Sergent et al., 2011, 2013; Thibault et al., 2016).

(iii) As delineated in §4.2.4 below, I take converging evidence, across a wide range of paradigms, to suggest that attention is a determining factor for how we temporally integrate sensory information. For example, manipulating exogenous and endogenous attention alters whether features are perceived as simultaneous (Donohue et al., 2015), and the timing and availability of attentional modulation shape whether features are correctly bound or misbound (Zivony & Emer, 2024). Neurophysiological work indicates that the temporal resolution of temporal binding is limited by attentional resources (Chennu et al., 2009; Correa et al., 2006). Even illusions such as colour–motion asynchrony can be eliminated simply by cueing attention to the relevant transients (Holcombe & Cavanagh, 2008).

(iv) Predictions and retrodiction alone cannot explain certain aspects of the flash-lag effect. These aspects may be better explained by appealing to some selection mechanism (like attention) that makes us prone to temporal errors (temporal misbinding) rather than motion extrapolation errors (Holcombe & Corbett, 2023).

To set out how attention might play this role in selecting which contents are temporally bound as an integrated perception, I will rely on a theory by Dołęga & Dewhurst (2020) that integrates a predictive processing theory of consciousness (Clark, 2013; Hohwy, 2012, 2013c; Seth, 2015) with Dennett's multiple draft model (Dennett, 1991).

### *4.2.1 The Winning Hypothesis Model of Consciousness*

In the predictive processing literature, it is often assumed that the predictive perceptual model that becomes 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, 2013c; Seth, 2015). This is the winning hypothesis theory (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).

Dołęga & Dewhurst (2020) argue that the winning hypothesis theory lacks a clear criterion for distinguishing between what makes the content of our brain's predictive models conscious and unconscious. They use the cases of unconscious representation and unconscious perception to show this. I 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 alternation between the 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 (both images are unobstructed and the perceiver is exposed to them in the same way), 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 subjective reports). But if one is committed to a strong representationalism (as implied by Hohwy, 2013, 184) according to which phenomenal experiences are determined by and identified with representational content, 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 the 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 content and phenomenal experience can come apart. This opens the possibility that one of the stimuli is always merely phenomenally consciously experienced, with no cognitively accessible representation being available to the person enjoying that experience. This aligns with Block's (1995, 2007) & Lamme's (2006) position that we have merely phenomenal conscious experiences of much more than what we have representational and access conscious experiences of. This is because the former depends on high-capacity early sensory processing, whereas the latter also depends on limited-capacity higher-order processing in the prefrontal cortex (Block, 2019). In some places Hohwy seems to adopt this distinction (Hohwy, 2012) and adapt it to his predictive processing 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 experience).

But elsewhere it is not entirely clear that Hohwy fully adopts 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’ (Jackson, 2003, 270), a claim that proponents of phenomenal consciousness straightforwardly deny: ‘The controversial part is that I take P-conscious [merely 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 content (Dołęga & Dewhurst, 2020).

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

The former option is untenable for multiple reasons. First, 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 is just a particular kind of processing rather than a specific stage of processing (Dołęga & Dewhurst, 2020). It is hard to see how 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 (Cohen et al., 2016; Cohen & Dennett, 2011; Naccache, 2018). This is because investigating conscious states requires eliciting some kind of report, and phenomenally conscious states do not allow for subjective reports of the kind that would allow one to infer anything about their contents. 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 forcefully undermined 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 such an account to give a clear description of how content and consciousness relate, in order to explain how contents are selected for conscious, temporally bound experiences.

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 our behaviour can be perceptually guided without involving any phenomenal experience. We see this in cases of blindsight (Azzopardi & Hock, 2011), where people who are phenomenologically blind (report not being able to see anything) still fare better than chance at certain visual discrimination tasks.

Since the perceptual hypotheses formed in blindsight or by motion-blind patients can affect their behavioural responses just as conscious states can, these perceptual hypotheses must be winning hypotheses at some level. Remember, according to the winning hypothesis theory, a perceptual hypothesis must be a winning hypothesis to determine behaviour.

Yet the winning perceptual hypothesis that guides behaviour is never phenomenally experienced. As such, the winning hypothesis account is not a theory of phenomenal consciousness, because to account for these behavioural effects, the winning hypothesis theory must allow 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 (insofar as there is a perception that influences behaviour just as a conscious state would) without this hypothesis being consciously accessible. Therefore, something more is required for representational content to become conscious.

### *4.2.2 Attention as the Difference Maker*

Dołęga and Dewhurst (2020) take this approach and argue 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, by combining Dennett's multiple drafts model (1991) with Hohwy's (2012, 2013) predictive processing theory. Once we have set out this account, I argue in §4.2.3 that we can straightforwardly apply it to the hybrid model of temporal binding to provide an account of which contents take part in our conscious temporally bound perceptions, and which do not.

To recap, the multiple-drafts theory says that for representational contents 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, i.e., how our attention is directed towards processed contents.

I will use the term attentional modulation to talk about the broader function of how information processing and the deployment of cognitive resources are altered. As such, I do not want to identify attention either with a resource or a mechanism. Rather, I take attentional modulation to be a wholesale way of talking about the information-processing mechanisms and resource management systems that play a role in subserving the functional role of 'mentally selecting content for behavioural purposes' (Wu, 2024). We can use this term to implement a more definite structure of how different drafts are probed at different times, i.e., which draft we are conscious of (Dołęga & Dewhurst, 2020).

More precisely, attentional modulation allows our perceptual system to selectively deploy resources and alter information-processing to change the kind of access that different mechanisms have to certain contents and incoming stimuli. As William James put it, attention 'implies withdrawal from some things in order to deal effectively with others' (James, 1890, 404). When the attention to some processed contents or incoming stimuli is modulated, the importance and trustworthiness of those contents and stimuli are boosted at the cost of ignoring other contents and stimuli. While attention is a multi-faceted phenomenon with no clear unified mechanistic or neurophysiological meaning (Anderson, 2011; Di Lollo, 2018; Hommel et al., 2019; Hommel & Colzato, 2015; Krauzlis et al., 2014; Mole, 2011), all the many different ways in which attention can be selective fit well with the broader functional description of being a 'mental selection of a target to guide behaviour' (Wu, 2024, 306)[14]. This construct is generally applied across almost all

---

[14] Attention has as a theoretical concept been subject to increasing scrutiny causing some researchers to mount sceptical arguments against its validity (Anderson, 2011; Di Lollo, 2018; Hommel et al., 2019; Hommel & Colzato, 2015; Krauzlis et al., 2014; Mole, 2011). Such discussion often concerns theories of attention as a whole, but explicit analytical attempts to divide attention into categories like endogenous and exogenous attention are also under fire. The issue is that not all phenomena fit neatly into these two categories. Sometimes, endogenous attention is too automatic, and exogenous attention too goal-driven (Awh et al., 2012; Folk et al., 1992; Hommel & Wiers, 2017). Awh et al. (2012) attempt to accommodate these cases by adding a third category of attention, selection history, which although working as intended it, according to Hommel et al. (2019) merely patches a whole rather than getting at an answer to what mechanisms drive attention. Moreover, the term 'attention' is criticized for being used to refer both to the problem and the solution in cognitive processing—serving as both the explanandum and the explanans (Hommel et al., 2019). All this being said, there is reason to assume that attention and the categories of endogenous and exogenous

paradigms empirically testing attention. In this chapter, I focus on just two popular notions of selective attention, endogenous and exogenous attention (however, many other kinds of attention would fit within the proposed framework).

Exogenous attention is commonly taken to be bottom-up, stimulus-driven, automatic and transient (Yantis, 2008). In this case, it means that exogenous attention can select for 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 becoming conscious. For example, a sudden noise or an ingrained, recognisable stimulus 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, selects for 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 example, if we in detection tasks, endogenously attend to a Waldo-like character, our perceptual system is more likely to respond when our eyes finally fixate on Waldo.

Both these ways of conceiving attention fit the widely endorsed functional role of attention, it being the ‘mental selection of a target to guide behaviour’ (Wu, 2024, 306). Put in a snappy way, endogenous attention is the mental selection of a target for behaviour set by goals, intentions, beliefs, bias, etc., while exogenous attention is the mental selection of a target for behaviour set by biased saliency maps responding to relevant salient stimuli in the environment (Wu, 2024).

Including an account of how attentional modulation plays a role in determining what contents become conscious also explains how otherwise low-probability, unexpected, or ambiguous stimuli can become conscious. In the case of low-probability stimuli, exogenous

---

attention can be used in good faith and live up to a range of functional and operationalization criteria. According to Wu (2024) it is possible to derive a very clear and common conception of attention that most researchers agree on, by simply looking at the design structure of the burgeoning experimental psychophysical, neurophysiological, and cognitive science literature on attention (>45.000 studies). Most are designed around a common conception of attention where attention is that which selects targets for certain actions by directing attention towards some things and withdrawing attention from other things. Attention allows us to deal effectively with a select few items. e.g., selecting spatial size and shape features to guide grasping. This provides a substantive causal account of the role attention plays (Wu, 2024). If we use this common conception of attention as the explanandum of cognitive science, neuroscience, and psychology research on attention, there is no reason to think of attention as also playing a role as explanans. This also goes for endogenous and exogenous attention.

attention may select that target if it is relevant for actions (e.g., avoiding an unexpected cyclist) or simply if the target is a particularly salient stimulus (Bowman et al., 2013). In the case of ambiguous stimuli, endogenous attention affects processing top-down by focusing on the processing of specific targets that fit with certain known object and action categories (see Çukur et al., 2013; Lupyan, 2015).

By combining the predictive processing framework with the multiple drafts model and a model of attention, we assume that a variety of contents exist at different hierarchical levels, and that they become conscious when attention probes them 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. Thus, a perceptual draft only counts as conscious when it is both highly probable and currently probed by attention[15]. Moreover, and importantly, the multiple drafts framework allows attentional modulation to take place retrodictively rather than always playing a predictive role. It is these extensions of the predictive processing framework that allow us to 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 (sometimes 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 for which one of the images remains dominant depends on how much it consumes attention (Zhang et al., 2011).

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)[16]. There is no cognitive access without any attentional modulation of the contents that causes them to grab control over the perceptual system (or so I assume here, I discuss this assumption critically in Chapter 5). Yet this does not prevent the given contents from

---

[15] It is an open but perhaps empirically intractable question whether a coalition of contents determine the content of consciousness before the moment at which a draft including these contents is probed (see Akins, 1996).

[16] 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).

constituting a good predictive hypothesis at some level in the hierarchy and does not prevent them from influencing behaviour subconsciously.

Moreover, this also explains a series of non-pathological cases where a near-threshold target stimulus that is normally not consciously perceived is more likely to be consciously perceived if attention is cued to the target's location up to 100–400 ms after the offset of the target (Kentridge, 2013; Rimsky-Robert et al., 2019b; Sergent et al., 2013; Thibault et al., 2016). The fact that people are more likely to perceive the target when attention is cued to the target location after its offset can be explained by the fact that the unconscious representation of the target becomes consciously available through being retrodictive attentionally modulated (rather than being predictively attentionally modulated).

As such, Dołęga & Dewhurst (2020) argue that we can better account for how content relates to conscious and unconscious perceptual experiences when we specify how attention influences which content becomes conscious.

### *4.2.3 Attention and Temporal Binding*

By using the framework set out above, we can quite briefly set out the second hypothesis: Attentional modulation determines which contents are integrated into a temporally bound conscious perception, by modulating both predictive and retrodictive processes.

This means that attentional modulation plays a selective role in temporally organising perceptions such that specific sensory contents are selected for guiding 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[17]. Attentional modulation allows perception to focus selectively on targets relevant for guiding behaviour in a way that is biased top-down by goals, intentions,

---

[17] By perceptual field I mean the area of external stimuli processed by, or at least immediately available to being processed by, our sensory system. It includes the array of sensory signals that our brain can contemporarily process across sensory feature types and modalities. I do not take the perceptual field to be necessarily all integrated sub-personal processing as such integration may require higher-order processing, and I want to reserve the perceptual field as referring to the array of sensory features processed by, or at least immediately available to being processed by, our sensory system. Moreover, I do not take the perceptual field to be isomorphic to our phenomenal field where the phenomenal field is all our concurrent conscious perceptual experiences (although some argue that these may be isomorphic (Block, 2007)). What it may take for features in our perceptual field to be consciously perceived, and thus part of our phenomenal field, will be discussed throughout this chapter.

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 features to bind. According to the second 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 to 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 video games, navigating traffic, hunting, and other activities that require accurately organising temporal information into synchrony and order relations, we need to be able to select which of an abundance of contents to temporally organise to engage successfully in that task. The evidence from sections 2.1 and 2.2 clearly shows 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 the flexibility of how our system engages in temporal binding, and how it is adaptive to all sorts of contexts. This flexibility and adaptability come from attention continuously selecting for different sensory contents in light of endogenous and exogenous bias. Due to this flexibility, what counts as 'the best temporal organisation of our perceptions' changes, and so do our conscious experiences of what features are perceived as temporally bound. Consider some of the cases we have already discussed.

We can explain the illusion where one sees a set of 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 enforces a retrospective redrafting of the other lines as being offset as well. The involvement of attention is clearly crucial, since the retrodictive effect of the offset does not affect unattended streams (Drissi-Daoudi et al., 2019). Similarly, the other cases of retrodictive effects on perception mentioned above (§2.1) rely on either bottom-up or top-down modulation of attention (Drissi-Daoudi et al., 2021; Rimsky-Robert et al., 2019; Sergent et al., 2011, 2013; Thibault et al., 2016).

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), can be explained as cases where endogenous attention to causal patterns or speech predictively influences how we temporally organise the relevant inputs.

There are also cases where it might be argued that exogenous attention causes a retrospective redrafting of a predictive draft. In the case of motion trajectories with subsequent unexpected motion reversal, we would predict that we 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 selection 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 hypothesis can be used to integrate a predictive event time theory with a retrodictive event time theory such as the content-sensitive settling account. With this 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 temporally bound perceptions based on a predictive hypothesis, or when the task demands, goals, context, or salience of stimuli change significantly in a way that draws attention, it may force our perceptual system to retrodictively redraft our perceptions, thus redrafting which contents are temporally bound with which.

### 4.3 Empirical Support for the Role of Attention in Temporal Binding

While there is only a small subset of experimental work that directly tests the role that attention plays in the integration of temporal information, I want to explicitly point out these, as they uniquely support the second hypothesis of the hybrid theory. That is, the hybrid theory is exceptionally well positioned for capturing the data of these studies in ways that other theories cannot, because other theories do not incorporate attention as a central mechanism of temporal binding. The following can, as such, be seen as supporting how the hybrid theory does not just provide an a priori story, but a theory useful for interpreting and guiding empirical work on temporal binding.

Donohue et al. (2015) show how the manipulation of exogenous spatial attention, perceptual goals, and task demands has 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.

In a new study explicitly focused on testing the role of attention in temporal binding tasks, Zivony & Emer (2024) find, contrary to Vul & Rich (2010) (who argue that attention is not necessary for temporal binding), that attentional modulation directly influences how features are bound together in time. Zivony & Emer (2024) show how the timing of attentional modulation plays a crucial role in determining whether co-occurring features are accurately perceived as part of the same object or are perceived as misbound.

Relatedly, Chennu et al. (2009) use 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 attention (which is limited or at least depends on limited resources, see Wu, 2024). Another study by Correa et al. (2006) investigates how paying selective endogenous attention to specific moments indicated by temporal cues positively affects the temporal resolution at which we can engage in temporal binding.

McDonald et al. (2005) review how we should interpret the well-known phenomenon (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). Traditionally, 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.

A study by Holcombe and Cavanagh (2008) (mentioned in Chapter 2, §2.1) 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 equally rapid access to the transients of both colour and motion. 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.

As also mentioned in §2.1, 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.

Two last potential theoretical benefits of the hybrid model come from ongoing research in how attention, action, and predictive perception relate. First, the model fits well with evidence of how predictions that contribute to visual tracking of objects only occur when we deploy attention to the relevant features (Howe & Holcombe, 2012; Hogendoorn, ACNS 2025), and why the neural pre-activation due to predictive processing only seems to occur when needed for some judgments or actions (Gastrell, ACNS 2025). This may be because attentional resources are required for us to be able to integrate temporal features of where and when objects are located, and attentional resources strongly favour stimuli that are relevant for behaviour, e.g., relevant to goals, intentions, actions, judgments.

Second, because of the attention component being central to how we temporally integrate sensory information, this is more open to a diverse set of explanations of what underlying mechanisms are involved in temporally integrating sensory features. Temporal binding is a complex phenomenon, and it is unlikely there is a single underlying mechanistic account for it. To give a concrete example, consider the case of the flash lag effect explained in §2.2. Some have argued that the perception of the illusory misalignment of the flash and the moving bar cannot be fully explained by appealing to prediction or retrodiction alone (Holcombe & Corbett, 2023). In fact, there exist many plausible accounts that explain the effect as being partially due to some temporal error in a perceptual mechanism. These alternative explanations are (a) *triggered position sampling:* the occurrence of the flash may trigger a delay in sampling the location of the moving object causing the perceived misalignment (Schlag & Schlag-Rey, 2002); and (b) *temporal misbinding:* the flash is mistakenly bound to a later position of the moving object either because we cannot consistently sample the position of the highly temporally uncertain flash (Shen et al., 2007) or because the moving bar is a second-order change that only produces a temporally uncertain transient (Nishida & Johnston, 2010)[18]. Both these accounts have to do with how and when attention samples information about the features to be temporally bound. By placing attention at the centre, the hybrid theory gains some explanatory flexibility which is necessary for a functional theory of temporal binding because temporal binding is likely affected by a host of underlying mechanisms rather than a single cause (as is suggested by the brain time theory (see Chapter 2) or simple predictive extrapolation accounts (as many defend due to the flash-lag effect, see Holcombe & Corbett, 2023).

## 5 Consequences for Time Perception

Before wrapping up this chapter, I want to present three interesting consequences for time perception that follow as a result of the hybrid theory. These are intended as very general and significant consequences related to the broader philosophical discussions of how our temporal experiences are structured.

---

[18] Note the theory is not compatible with the explanation that Holcombe & Corbett (2023) call intermittent position updating. On this account the flash lag effect is due to processing being discrete resulting in the location of the moving object not being perceptually updated until the object and the unmoving flash share a later concurrent discrete temporal window (Chakravarthi & Vanrullen, 2012; Chota & VanRullen, 2019)). This theory is incompatible with the hybrid theory because it rests on a theoretical framework according to which conscious perception is a discrete process that takes conscious perceptual updating to be periodic, i.e., conscious perception is not a highly flexible and adaptive phenomenon nor is the time of a conscious perceptions a temporally indeterminate as the hybrid theory suggests.

### 5.1 Consequence 1: Conscious Experiences Are Temporally Indeterminate

According to the hybrid theory, there is no reason to assume that there exists any independently identifiable onset or offset of consciousness. To quote Dennett & Kinsbourne (1992, 237): '[T]he discovery of a "now" token in the brain hopelessly underspecifies a time of intended "utterance".' The concurrent neural events that correspond to some alleged 'now' token will, on the hybrid theory, potentially be spread across a period of predictive drafting and retrodictive redrafting, with no way of specifying a determinate moment at which a conscious 'now' occurs.

Given that the content of our temporally bound perceptions depends on attention, there will be no probe-independent temporally bound perceptions. Consequently, the way in which we are probed necessarily influences the content of our reports. Reports therefore do not provide momentary slices of information about the time of representing, but only information about the time represented. The moment at which we report our perception of temporal binding does not reveal when the relevant conscious experience occurred.

Our conscious experiences are temporally spread out because they are continuously undergoing predictive and retrodictive revision. At best, a response (verbal or non-verbal) to some task or other probe, determines only the time represented in experience, not the time at which that experience was formed.

Neither the subject being probed nor the person probing the subject has privileged access to when the experience underlying a behavioural response was formed or revised, or whether it was based on a retrospectively contaminated memory of some experience.

Probing someone mid-processing may be useful for discovering how early certain sensory discriminations are achieved in a processing stream, but this comes 'at the cost of disrupting the normal progression of the stream' (Dennett, 1991, 136). Probing will always influence temporal binding, given that what is temporally bound depends on attentional modulation.

Ultimately, this implies that any sharp distinctions between whether a temporally bound perception is due to current processing of temporal relations, recent memory of temporal relations, or prediction of temporal relations break down. None of these contributes to a more direct or privileged experience of the present than the others. All contribute to perceptual experience, temporal judgment, and time-sensitive behaviour.

For these reasons, we should avoid conceiving of conscious experience as a series of temporally determinate events structured by the order in which information crosses a fixed boundary between unconsciousness and consciousness. Instead, we should understand conscious experiences of temporal binding as temporally indeterminate, continuously shaped by predictive and retrospective processes, and influenced by ever-changing bottom-up and top-down factors. There is no 'turnstile of consciousness' by which features line up in neatly ordered groupings before entering awareness.

To flesh out this idea, I provide an instructive and extensive metaphor below of what it would mean for conscious experience to be temporally indeterminate in this way. This metaphor is an amalgam partly inspired by Dennett (1991), Dennett & Kinsbourne (1992), Hogendoorn (2022), and Prosser (2016). Here goes:

When we think about a large organisation such as the BBC, whose main purpose is to constantly and continuously broadcast news about the world, we should not think that the order in which news is broadcast by the BBC is determined by the order in which information enters the BBC. This is because it is very hard to say exactly when information enters the BBC. Many departments in the BBC receive information that is integral to the process of broadcasting news, and information from multiple departments often needs to coordinate and integrate their information to come up with good, coherent news stories that can be broadcast in a way such that the news stories more or less accurately fit with the facts in the world that they are meant to represent.

When a news story has been drafted by integrating the different pieces of information from the different departments and is broadcast, the news story is informative about the world at the time it is broadcast (it is up to date) and gives people a fairly accurate sense of what information the BBC has tried to represent (it is accurate). However, the time at which the news stories are broadcast and the order in which the news stories are told do not necessarily reveal anything significant about the order in which the BBC has received the different pieces of information that are cobbled together to form this series of stories.

It is very likely that some departments in the BBC try to predict aspects of the story that the department has not yet received information about from the world by fitting current expectations, context, and representational goals to the information already available. This predicted information may play an important part in what is broadcast, even though it precedes information that any department has yet to receive from the world. After all, the BBC must constantly keep broadcasting and thus cannot always wait until all the data has come in (also, how would one even know when all the data is in?). Predicting the news well is also fruitful when you get it right, because it helps the BBC one-up other broadcasting companies that are worse predictors, and beat those who do not engage in prediction to the scoop.

But of course, the BBC is bound to get something wrong, and predicting can be a dangerous game, since wrong predictions would result in bad consequences (inaccurate news may lead to a bad reputation for providing untrustworthy news). To avoid this, the BBC may continuously redraft its broadcasted news to fit new information and new predictions from different departments of the BBC, which is crucial for making the story a more accurate representation of the facts of the world it is trying to represent. This redrafting might happen so smoothly and quickly that no one notices or remembers that the story used to be different and has gone through multiple drafts. Maybe the BBC broadcasting system itself does not even notice that this is what is happening because the departments are semi-disconnected from each other, and it all needs to run quickly and smoothly without anyone in particular overlooking all the drafts and redrafts (maybe it is even redrafted before the other version gets to be broadcast). Broadcasting itself is a temporally extended event, so redrafts may also take place on the fly.

In this conglomerate, the BBC news must both be as up to date as possible and as accurate as possible. While these two qualities, speed and accuracy, compete with each other, the BBC can outsource the process of drafting to a large range of specialised departments that all get information from the world, predict information to fill gaps, and redraft these predictions as better predictions or new information comes in, to meet both these qualities as well as possible.

A last important thing to note in this metaphor is that the BBC can only ever broadcast one story at a time, despite having information available in different departments that could be used to draft and broadcast other stories. This is because the parallel information gathering being done by the many departments vastly overflows the serial broadcasting capacity. The BBC must choose one story to broadcast and neglect others, and this may depend on a lot of factors, both top-down and bottom-up. Internally, there are bottom-up criteria about the kind of information that is coming into the departments (that may draw more or less attention) and top-down criteria regarding how well the information of different departments fits together for drafting accurate and coherent stories (departmental editors play a role here). There are also more global top-down criteria for what best meets the demands of the public and fits with news that has just been broadcast moments before. Due to news having to be news, it may also be that some stories never get broadcast because they did not satisfy the different criteria as well as other news at times, and because they now no longer live up to being news (they are old).

The idea is that, beyond being subject to speed and accuracy constraints, there is also a ‘one story at a time’ criterion, which forces the BBC as a whole to ‘select’ what should be broadcast (the information in the BBC vastly outstrips what could potentially be broadcast). There is, however, no central editor-in-chief who alone is in charge of this selection, nor is the selection fully determined by the order in which information enters the BBC. The selection is also partly settled by all manner of internal as well as top-down and bottom-up criteria.

Importantly, the only real clue to what information is available when to the BBC is the broadcasts themselves or a transcript thereof, and below a certain timescale, this does not reveal anything determinate about the temporal order in which information has entered the BBC or the time it became part of a broadcast.

While this is a mere metaphor, the hybrid theory suggests that our perceptual system works a bit like this, where the broadcasts are our reported/reportable experiences, which are affected both by the bottom-up flow of information from the world occurring in parallel and the top-down flow of information from higher-order predictions, expectancies, beliefs, perceptual goals, etc.

If this is how our temporal perceptual system works, then there is no temporally determinate fact about when, in the information processing story, an experience occurs. While it is possible to ask a person about their experience at a specific time, it is not clear that doing this is tantamount to making a cross-section in the brain revealing the current moment of perceptual information processing, such that the reports give us an indication of what content is currently inside the boundaries of conscious states and in what order it crossed this boundary. This is just like how watching a news story being broadcast from the BBC at some determinate moment does not provide us with evidence of what information is currently available at the BBC at that specific moment. Some selection has taken place, and this selection is not only determined by when the information comes into different departments of the BBC. Some information is predicted before that, and some information is redrafted retrospectively, and all is done according to a series of external and internal criteria.

While the idea that our conscious experiences do not have any temporally determinate onset or offset below some timescale might seem contrary to our experience, it is still compatible with the phenomenology that every conscious event seems temporally determinate, or at least these perceptions seem to be presented to us as having perfectly temporally determinate contents, and these perceptions seem to be presented within a single unified temporally determinate timeline. In Chapter 4, I propose an explanation for why we seem to experience perceptions as contained within a temporally determinate ‘now’ and as making up a determinate unified timeline, although the onset and offset of conscious experiences are temporally indeterminate.

### 5.2 Consequence 2: Temporal Binding is Likely a Necessarily Conscious Phenomenon

This follows, albeit not directly, from the idea that our temporally bound perceptions depend on the attentional modulation of representational contents in the relevant way to be temporally bound.

By this I do not mean to claim that attention is identical with consciousness, nor that any attentional modulation is sufficient for consciousness. There are many studies showing that this is not true, e.g., spatial attention can contribute to discrimination tasks without resulting in a conscious experience of the attended features (Kentridge, 2004), attended features can contribute to priming effects while remaining consciously inaccessible (Cohen et al., 2012; Dehaene & Changeux, 2011; Naccache et al., 2002), attention to semantic information can subliminally influence success on forced-choice paradigms (Dehaene, 2014, p. 201).

Rather, within the hybrid model, attentional modulation is the mechanism by which a perceptual hypothesis gains system-level control. A perceptual draft becomes conscious when it is both sufficiently supported within the predictive hierarchy and probed by endogenous or exogenous attention, thereby gaining access to higher-order bottlenecked processes involved in reporting, deliberation, and action guidance. Contents that are processed but not attentionally modulated may influence behaviour in limited, automatic ways, but they do not achieve the kind of global accessibility characteristic of conscious perception. In this sense, attentional modulation functions as the difference-maker between unconscious predictive processing and consciously accessible perceptual experience.

However, because temporal binding is a selective phenomenon in this way, we do not temporally bind all processed sensory features. There will be many sensory features that, while processed, remain unconscious influences on our sensory-motor system. Not every action needs to be consciously mediated. Unconscious contents that are not attended to might still influence our behaviour in different ways, but there is nothing that requires these contents to be temporally bound. In fact, according to the hybrid theory, it is likely that they cannot be temporally bound and remain unconscious. This means that our neural processing systems can carry information that unconsciously influences our temporally coordinated behaviour without necessarily implying any larger temporal integration of this information across these systems. Only when contents are attentionally modulated top-down or bottom-up in the right way can they become part of a temporally bound conscious experience.

At present, there is no clear-cut empirical data or research paradigm that could provide definitive and uncontested support for this. In fact, the answer is underdetermined by the data, and settling anything on the basis of what we know now risks begging the question against certain theories. This is due to a long and rich debate about what it means for visual perceptions (as well as any other perceptions) to count as conscious rather than unconscious, as briefly discussed in §4.2 (and discussed extensively in Chapter 5). For example, would we take something to count as a conscious visual perception just in the case that our actions and reports are guided in ways that rely on visual perceptual processing mechanisms, even if the subject reports that they did not have any visual phenomenology of said perceptual experience?

This debate often centres on how to interpret a series of case studies. One important case study is the study of blindsight mentioned briefly above. The term ‘blindsight patients’ is used as an umbrella term to cover several kinds of blindsight patients who all have a significant lack of visual phenomenology. The lack of visual phenomenology is described across three dimensions: ‘subjective sensation’ (i.e., feeling something), ‘objective capacities’ (i.e., discriminating the form of a stimulus), and ‘visual’ sensation (i.e., seeing the stimulus). By categorising blindsight patients across these dimensions, one can set out six distinct categories: complete hemianopia, blindsense and ‘vision’ blindsense, type 1 blindsight, type 2 blindsight, and ‘visual’ sensation (Derrien et al., 2022). While blindsight patients differ across these three dimensions, the relevant common factor for our purposes is that these people are phenomenally blind to most visual features, if not all, visual features of objects, despite the fact that they can (only when probed) make better-than-chance guesses in visual tasks (such as detecting a moving stimulus in their scotoma). This is empirically verified and uncontroversial. Yet controversy arises when we have to interpret whether or not blindsight patients are conscious of any of these visual properties in their scotoma that they can make better-than-chance accurate judgments about. Some might argue that blindsight patients have a degraded conscious visual perceptual experience of the stimulus (Mazzi et al., 2016; Overgaard et al., 2008; Phillips, 2021; Sahrie et al., 2010), while others will claim that blindsight patients ‘merely’ have a functional non-conscious perception of the stimulus (Azzopardi & Cowey, 2001; Azzopardi & Hock, 2011; Persaud & Cowey, 2008).

However, saying one or the other depends on (a) the details about the degree of phenomenology, functional competence, and visual sensation of different blindsight patients (for further discussions see Derrien et al., 2022; Michel & Lau, 2021; Peters et al., 2016) and (b) the theory of consciousness one ascribes to. Some theories of consciousness would take this degraded perceptual awareness to count as a kind of conscious perceptual experience, regardless of whether we can report it or not, as long as it depends on the right kind of early sensory recurrent visual processing loops (Block, 2007, 2011; Lamme, 2001, 2006; Lamme, 2010). Other theories would argue that this kind of degraded residual processing connectivity cannot generate the required kind of sustained activity needed for higher-order processing (Lau & Rosenthal, 2011) or global connectivity with prefrontal areas of the brain (Dehaene, 2014) required for conscious perceptual access. As such, an answer to how consciousness and perceptual capacities relate (in the context of visual perception) is empirically underdetermined and risks begging the question.

The same goes for the relation between consciousness and our capacities to engage in temporal binding. Would we say that a person who reports that they have no phenomenal experience of perceiving any features as temporally bound, yet who can perform better-than-chance on simple temporal binding tasks, is having a conscious temporally bound perception? Or must the person in question describe themselves as having a phenomenology of temporal binding for it to count as a conscious temporally bound perception?

Despite these methodological issues, one might think that testing the temporal binding capacities of blindsight patients might be instructive to see how temporal binding performance and conscious experience relate. This may also be particularly interesting because it has been suggested that evidence from blindsight studies (such as Kentridge, 1999, 2004) and semantic priming studies (Neumann & Klotz, 1994; Marcel, 1983), shows how spatial binding of features can occur unconsciously (Bayne, 2010, 231). Though because of how spatial binding is likely to differ from temporal binding (see Chapter 1) we cannot simply assume that this is the case for temporal binding. Spatial attention allows for the faster selective cognitive processing of certain features in feature maps that can then be integrated in early visual processing, however, as I have argued, it is unlikely that temporal binding is the consequence of synchrony or other kinds of mappings in early cortical processing. I have suggested that something more is needed for temporal binding. While I have suggested that this means that temporal binding is likely to be necessarily conscious, these considerations are based on theoretical and explanatory claims. We need separate studies on blindsight and temporal binding in order to assess whether unconscious temporal binding would be possible at all.

A blindsight patient might be able to temporally bind two distinct features in their phenomenally blind perceptual field (e.g., whether two motion trajectories switch direction at the same time or not). Or it may be that blindsight patients can temporally bind an auditory feature (of which they are phenomenally aware) with a phenomenally blind but 'perceived' visual feature. This would lend some credibility to the idea that temporal binding of visual features is possible without visual phenomenal awareness.

However, it is not clear what kinds of results we should expect. It is unlikely that they would be able to perform above chance on temporal binding tasks because it would require the person to be able to say something about the temporal relation between two stimuli. This requires some kind of semi-determinate temporal representation of when the two occurred relative to each other, and blindsight perceptions are not determinate but are weak, uncertain, and vague. The detection of the presence of a change in luminance that is presented across 2 seconds, which it takes for blindsight patients to discriminate change, is hardly likely to underpin a synchrony judgment (Sahraie et al., 2010).

Instead of relying on the study of blindsight, one might use a subliminal processing paradigm to test whether we can temporally bind features without the involvement of any conscious experience. Subliminal processing refers to a non-conscious type of neural processing where the processed information remains consciously inaccessible because it does not carry the right kind of bottom-up activation for conscious access (Dehaene et al., 2006). Subliminally processed information will be completely invisible to the perceiver, regardless of the attention and effort exerted (a subliminal stimulus will never become conscious).

A subliminal experimental paradigm is then one in which a person is tasked to view a series of images, some of which are subliminally processed. The subliminal processing of images can be forced by flashing the target image very briefly (often below 50 ms) and sandwiching it between two other images that function as a mask of the target image (Dehaene, 2014). It is not the short duration of the target image itself that makes it invisible to the perceiver; it is the visual 'masking' stimulus that makes it so that the perceiver only sees the masks. One can then test if the subliminally processed target stimulus influences the perceiver's behaviour by giving them a forced-choice task that can be performed at above-chance success only if the subliminal processing had a direct or priming effect (Cohen et al., 2012; Dehaene & Changeux, 2011; Naccache et al., 2002).

It may be possible to devise such a subliminal processing experiment for temporal binding. We could imagine an experiment where a person is tasked to temporally bind a colour with a motion direction. The person would be shown a series of very brief moving videos: one of a red pattern moving leftward and another of a green pattern moving rightward, both of which are masked. The task of the person is then to judge either the motion direction of the red pattern, the motion direction of the green pattern, the colour of the leftward-moving pattern, or the colour of the rightward-moving pattern. The response would be a binary forced-choice task that, depending on the scenario, would ask the participant to choose between left/right or red/green.

If a person can perform at better-than-chance rates on a subliminally processed temporal binding task, then this might indicate that some degree of temporal binding can take place in the absence of any phenomenal awareness of temporal binding. However, since no such studies have been undertaken, it would be entirely speculative to say anything about whether or not this is possible. Even if it turns out that subliminal temporal binding is possible, there are still reasons to doubt what such results say about the consciousness status of perceptual information in question. The interpretation of subliminal perception has historically been very disputed in psychology (Eriksen, 1960; Holender, 1986; Merikle & Daneman, 1998; Velmans, 1991) and is still unresolved (Dehaene, 2014; Faivre et al., 2017; Railo et al., 2021). There is neither empirical data nor any theoretical agreement that would allow us to definitively say whether we can engage in temporal binding without having a conscious temporally bound perceptual experience of the features in question.

### 5.3 Consequence 3: Temporal binding is Constrained by Attention

If the hybrid theory is correct that, for content to be consciously perceived as temporally bound, it must be endogenously or exogenously modulated, then, at any given time, a lot of sensory features within our perceptual field would not be consciously temporally bound (as per 'Consequence 2'). This is due to the large discrepancy between the capacity of sensory processing and the capacity of attentional modulation (Baars, 1993; Cohen et al., 2016). Hence, putting attention at the centre of temporal binding implies that there are attentional limits on temporal binding. This is a hypothesis that must be put to the test. In Chapter 5, I explore how this could be done by examining how our temporal binding capacities and our phenomenology of temporal binding might be (possibly differentially) affected when attention is distracted or overloaded. If the hybrid theory is correct, we should perform worse (or even fail) on simple temporal binding tasks when attentional resources are overloaded or distracted, or when attentional access to certain representational contents is restricted. This claim might seem uncontroversial and true, but when it is compared to how attention restricts our phenomenology of temporal experiences, the claim is more radical.

Pre-theoretically, our perceptual experience appears temporally bound in such a way that we do not ordinarily encounter parts of it as temporally disjoint or unbound. In Bayne's (2010) terminology, conscious experience exhibits 'phenomenal unity'. That is, the different parts of our experiences (different sensory features, regions, properties) are experienced together as having a single, conjoint experiential character. One way to interpret this phenomenal unity, in the temporal domain, is that perceptual features are experienced as temporally integrated within a single unified experiential moment. It seems very appealing to describe this phenomenology of temporal binding as a global perceptual phenomenon, meaning that all parts of our perceptual field are experienced as one temporally bound whole. This is also because it is not clear what the phenomenology would be where some experienced sensory features are not experienced as standing in a temporal relation with all other perceptually experienced features.

Yet, if the hybrid theory is correct, then our experience of temporal binding should be constrained by our attentional capacities and thus should not extend across our entire perceptual field, because it is impossible to fully attend to our entire perceptual field (Cohen et al., 2016). If the hybrid theory is correct, then there is no distinction between the phenomenal unity of temporal binding and the attended representational content that we are temporally binding at any moment. Consequently, 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. We explore this hypothesis in detail in Chapter 5.

## 6 Conclusion

Above I have described and set out the hybrid theory in terms of two hypotheses:

> ***Hypothesis 1:*** Our perceptual system uses retrodictive and predictive processing in mutually beneficial ways to engage in flexible temporally binding of sensory features.
>
> ***Hypothesis 2:*** Attentional modulation determines which sensory features become part of a conscious temporally bound perception by modulating said contents. This makes temporal binding adaptive to salient temporal patterns, goals, intentions, actions, and expectations.

This hybrid theory is a functional theory of temporal binding that describes how the brain integrates sensory features into a unified temporal experience via a hybrid process that combines predictive and retrodictive mechanisms with an account of how attention is central for determining which features are consciously perceived as temporally bound. On this view, available temporal predictions contribute to the rapid temporal binding of features so our perception may approximate the event time of these features, while retrodictive processing contributes by allowing incoming sensory information, changes in task demands, goals, context, or salience of stimuli that draw attention and clash with attended predictions to retrodictively redraft our perceptions, thus redrafting which contents are temporally bound with which.

As such, the hybrid theory defines temporal binding as a flexible and adaptive process. On this account, temporal binding is adaptive because the efficiency of temporal binding is dependent on contextual demands in a way that seeks to maximise ecologically useful information gained from temporal binding at minimal costs. Moreover, temporal binding is flexible because what features are temporally integrated (and whether or not they are temporally integrated) depends on what the situation calls for. By defining temporal binding as an adaptive and flexible process, the hybrid theory can address Problems (i) and (ii) set out in Chapter 1. Problem (i) is to give a functional account of the many effects that influence temporal binding, and Problem (ii) is to provide a solution to the problem of neural delay and the problem of desynchronisation. The hybrid theory addresses both in the same way. By deploying retrodictive and predictive processing in a way that is selectively dependent on attention-relevant factors, the theory can explain how desynchronisation in sensory processing can take place without always affecting which sensory features are temporally bound, and how we sometimes perceive stimuli as temporally bound faster than direct processing allows, mitigating neural delays. This trade-off between how accurately sensory features are temporally bound and how quickly temporally bound features are perceptually available is adaptive and flexible to the situation at hand, and as such, we should expect to see a lot of variability in temporal binding (addressing Problem (i)). Moreover, we should expect that desynchronisation or neural delays can be mitigated and compensated for (addressing Problem (ii)).

Note that this view does not state that the proposed interactions between attention, retrospective processing, and predictive processing are the only contributors to temporal binding, and thus that this theory tells the final story. The theory simply states that these factors and their interactions play a large role in temporal binding. That is, an account like the hybrid theory allows us to explain a range of effects on temporal binding as well as explain how our brain can compensate for neural delays and desynchronisation. It is on these terms that this theory is to be preferred over the brain time and other event time theories that leave unexplained things that the hybrid theory explains. In Chapters 4 and 5, I explore to what extent the hybrid theory can address the third and last problem, which involves ‘explicating the relation between the mechanisms underlying temporal binding and the content of our experiences of temporal binding.’

# Chapter 4

## Deflating Our Perceived Unity of Time

### 1 Introduction

Our perceptual experiences of time appear unified both synchronically and diachronically. As Viera (2020, 638) puts it, ‘While we perceive events in our environment through multiple sensory systems, we nevertheless perceive all of these events as occupying a single unified timeline. Time, as we perceive it, is unified.’ However, the mechanisms underlying time perception are both *fragmented* (time perception relies on many distinct sets of mechanisms distributed across the nervous system) and *dissociable* (the mechanisms can be intervened upon independently in ways that differentially influence our time perception). A theory of temporal binding should explain how our perceptual experience of time is connected to this underlying fragmentation and dissociation. Gerardo Viera has set out this explanatory problem succinctly:

> ‘Temporal perception is initially fragmented. The perceived unity of time is constructed from these fragmented capacities. However, at present, no theory explains this unity. Something new is needed’ (Viera, 2020, 655).

In this chapter, I provide a detailed phenomenological and functional description of three distinct aspects of this perceived unity of time and highlight why these are difficult to account for when such an account also has to connect this perceived unity to the psychophysical and neurophysiological data on how the mechanisms underlying time perception are organised.

Viera (2020) has argued that existing models of time perception, such as mirroring accounts (Dainton, 2000; Engel & Singer, 2001; Foster, 1991; Mellor, 1981; Miller, 1984; Phillips, 2014b; Rashbrook, 2013a; Zeki, 2007) and internal clock models (Chen & Yeh, 2009; Church, 1984; Gibbon et al., 1984; Hsieh et al., 2011; Treisman, 1963), fail to provide an empirically and conceptually convincing story that connects the perceived unity of time with the mechanisms underlying time perception.

In short, these existing models take the representational content of our experience of time perception to be determined by the structural organisation of the representational vehicles responsible for our time perception. Internal clock models treat temporal experience as the output of a centralised or functionally unified timing mechanism. Mirroring models maintain that experiential unity reflects synchronous or unified processing. However, following Viera (2020; 2019), I maintain that by grounding this unified experience in the representational vehicles, these models either idealise the structural organisation of temporal processing beyond what the data support or they fail to account for our perceived unity of time.

I argue that it is possible to provide a clear story about how the perceived unity of time relates to the mechanisms underlying time perception if we adopt the hybrid theory of temporal binding presented in Chapter 3. This theory accounts for how sensory features are perceptually temporally bound without this temporal binding being fully determined by the local processing of the underlying mechanisms. Instead, our experiences of time are also partly determined by the temporal organisation of representational contents taking place at a higher-order level, with attention playing a central role. A consequence of this theory is, however, that our phenomenology of the perceived unity of time is constrained by attention. I describe how this leads to a series of deflations of the phenomenological richness of this unity.

In §2, I describe the phenomenology and functions of the three kinds of perceived unity of time. In §3, I survey and present empirical data on the fragmented and dissociable organisation of the mechanisms underlying time perception and show how this complicates accounting for the perceived unity of time. In §4, I set out two standard accounts of time perception, the internal clock model (§4.1) and the mirroring model (§4.2) and examine how they aim to account for these different phenomena of our unity of time, yet fall short of addressing Viera's challenge, leaving either the perceived unity of time or the empirical data unexplained. In §4.3, I present an alternative way to approach Viera's challenge that fits with my hybrid theory of temporal binding set out in Chapter 3. In §5, I recap the central aspects of the hybrid theory proposed in Chapter 3, and in §6, I use the hybrid theory to account for the three perceived unities of time, thus meeting the explanatory challenge set out by Viera. I conclude in §7.

## 2 Three Different Kinds of Perceived Unity of Time

Viera (2020) describes three different ways in which our perception of time is unified and takes these as falling under the terms localisation and comparability, where localisation covers two quite distinct kinds of experiences. To avoid confusion, given that these three experiential phenomena are distinct, I refer to the three kinds of perceived unity of time using the terms temporal binding unity, timeline unity, and comparability. I describe these three phenomena in turn in §2.1, §2.2, and §2.3 by supplying a phenomenological and functional description of each. This will serve as motivation for why we should be concerned with accounting for these phenomena. In §3, I introduce the empirical data on how the mechanisms underlying time perception dissociate across a range of dimensions. Taken together, these sections make up a significant aspect of the explanatory challenge of explicating how the mechanisms underlying temporal binding relate to the phenomenology and content of our perception of time.

### 2.1 Temporal Binding Unity

*Temporal binding unity* can be phenomenologically understood as how perceived features seem to be temporally bound in time relative to a specific anchored moment, the moment we seem to live in at every waking moment. Experiences of temporal binding unity encompass all the experienced features that we take to occur synchronously with each other, or in other words, all the features that we temporally bind. Consider a flash, a sound, and a touch on your knee, all experienced as a synchronous, coherent perception. Phenomenologically, temporal binding unity is experienced not as a collection of separate, localized events, but as a globally unified moment in which perceptual features across the entire perceptual field from different modalities (colours, sounds, smells, and touch) are presented as unfolding together within a shared temporal frame. Rather than merely co-occurring, these features are experienced as temporally integrated, with no gaps or free-floating elements, and as jointly anchored within a single moment. A moment often denoted as our subjective 'now' (Callender, 2017).

Temporal binding unity also plays an important perceptual function. To effectively interact with our environment, we must organise the timing of features relative to one another (Holcombe, 2015; White, 2020). Crucially, we must also consciously represent when these features occur in relation to the present moment, our now (Callender, 2017; Hogendoorn, 2022; White, 2018). Coordinated behaviour depends on accurately estimating how external events unfold relative to the moment at which we engage with them. Simply perceiving that an event occurs is insufficient. We must also know when it occurs relative to our own position in time. Temporal binding therefore provides a temporal locative function that enables us to situate ourselves within the temporal structure of the world. In Chapter 1, I outlined five functions that temporal binding supports across perception, cognition, and action. These included:

(i) Maintaining a coherent and dynamically stable representation of the world,

(ii) Integrating features, objects, events, and contextual information within perception (White, 2020)

(iii) Binding higher-level semantic information such as emotional expressions (Kiss & Eimer, 2008), motion information of unobservable vectors (Hassenstein & Reichardt, 1956), and kinetic variables such as force, mass, causality, trajectories (Hogendoorn, 2020; White, 2009, 2012) with concurrent stimuli.

(iv) Providing a temporally structured input that can be used by other processes to support cognitive deliberation and action guidance (White, 2020).

(v) Allowing for a coherent and dynamically stable perceptual representation of the world across multiple timescales (Akyürek, 2025; Hogendoorn, 2022; Mégevand et al., 2013).

### 2.2 Timeline Unity

*Timeline unity* can be phenomenologically understood as how we can experience features of different modalities and occurring at distinct timescales, as taking place in the recent past and immediate expected future, all within the same unified timeline. We can both perceive features that change at the level of milliseconds, such as the onset and offset of a flashing light at intervals of 20 ms, and experience a long monotonous tone for 2 seconds, and experience these as occurring within a single unified timeline.

We do not experience features of different modalities (sight and sound) or features changing at different timescales (every 20 ms and every 2 s) as occurring within distinct timelines, which we then need to post-perceptually infer as occurring within some unified timeline. We simply appear to experience distinct features and changes occurring at distinct timescales as if they occur within one unified experienced timeline. This phenomenon of experiencing features (of different modalities, timescales, etc.) as temporally related (as occurring before or after) within one unified timeline is what we will refer to as timeline unity.

The experience of timeline unity also plays a functional role. Coordination of our actions may require that we consciously experience sensory features as within a unified timeline. This is supported by studies on working memory that show how we simultaneously maintain and diachronically integrate multiple perceptual features occurring across time to solve tasks that require representing the temporal order, direction, or other temporal relations between perceived sensory features (Baddeley, 1993; Ninokura et al., 2004; Roberts et al., 2018).

For example, in the case of speech comprehension, we must be able to maintain a small amount of phonological information for a short period to understand speech (Buchsbaum, 2013). This short-term conscious representation of phonological information has been described as a working memory capacity called the ‘phonological loop’ (Baddeley, 1993). Moreover, information from another modality, such as visual information of lip movements, can be added to the same timeline representation within which our phonological information is represented. This visual information is integrated with the auditory phenomenological information and maintained for a brief period (Tye-Murray et al., 2016).

The visual unified timeline and the phonological unified timeline of maintained information are not experientially distinct, even though they are processed by distinct mechanisms and are most likely never integrated by any third mechanism (Tye-Murray et al., 2016)[19]. We only seem to experience one unified sequence of events at a time, and while it is possible to imagine or represent multiple and dissociated timelines, it is hard, if not impossible, to imagine what it would be like to perceptually experience distinct features or events as ordered within multiple different timelines at the same time.

### 2.3 Comparability

The last phenomenon, comparability, can be phenomenologically described as the way in which the temporal relations of sensory features from different modalities, timescales, and other dissociating dimensions (§3) are experienced as belonging to the same kind of representational property, without requiring any active representational comparison. This kind of unity is importantly distinct from timeline unity. Timeline unity concerns the ordering of events within a single linearly temporal structure but does not by itself guarantee that temporal properties are represented in a common currency. In principle, the temporal properties of sensory features originating from different sensory modalities could be represented within a unified temporal order, while the temporal properties (durations or temporal locations) could be encoded in

---

[19] This claim is supported by other examples in auditory processing where distinct processing of individual frequency bands (Fagelson & Champlin, 1997) that are clearly experienced as a unified percept, do not seem to integrate at any distinct processing stage (Tye-Murray et al., 2016).

modality-specific formats that require translation. Comparability concerns the absence of such a translation requirement.

Comparing perceptually experienced temporal aspects (e.g., duration and time of occurrence) of auditory and visual sensory features does not appear to require any mental effort. Phenomenologically, there is no experience of time that is modally unique to, for example, hearing or seeing. That is, there is no distinct ear-time that is experientially different from eye-time. If you hear a sequence of sounds and see a sequence of flashes that are equally spaced out in time, the temporal experience of these two sequences is an experience of the same kind of properties.

Compare this to how the spatial content of different modalities is perceived in slightly distinct ways. There are experiential differences between spatial tactile properties and spatial visual properties. Consider Viera's (2020) example of comparing the intensity of a tactile sensation to the intensity of a light. Such cross-modal comparisons are possible, but as Viera (2020) highlights, there is a 'lingering awkwardness' to them insofar as the intensities do not seem to be of the same type of property. There are two different kinds of intensities in play. This awkwardness of cross-modal comparisons is absent for perceptual comparisons in the case of time (e.g., comparing the duration of a touch and the duration of a flashing light or comparing the temporal order of different modality pairings). Although seeing and hearing temporal duration or order are two different ways of representing time, just as seeing and hearing intensity are two different ways of representing intensity, the content of the durations is not experienced as two different kinds of content but as one kind of content.

While I agree with Viera, the example may be slightly misleading as the difference in comparability could arise because something like intensity does not have any naturally defined commonalities across modalities, i.e., there is no common metric according to which we can compare different modal representations of intensity. Instead, consider an analogy where such commonalities do exist. Consider the case of comparing spatial positions or distances across modalities. Comparisons of spatial position or spatial distance have natural commonalities across modalities. If I see a light 2 meters away or hear a sound that is two meters away, the two modalities represent the same kind of content: spatial distance. The same goes for spatial position.

However, the intuitive awkwardness of having to compare the temporal and spatial experiences remains. There is some effort involved in comparing the distance of a flash to the distance of a sound. Maybe this is because there are differences in the kinds of inferences and processes one utilises to judge distance via sound and vision (Anderson & Zahorik, 2014). The same goes for spatial position. Let's say you feel a touch on your back and have to report your representation of the spatial location of that touch. Afterwards, a bright dot of light is flashed on your back, and you have to turn your head (or use a mirror) and report your representation of the spatial location of the dot of light. When we are comparing the position of the touch to the position of the light, there is an active comparison that we need to perform. We need to consider the representations and compare their content. This need for active comparison does not seem to be the same in the case of cross-modal comparisons of the timing or duration of modally distinct features.

We can illustrate the phenomenology of comparability by providing a temporal analogue of Molyneux's question, to see how these evoke different intuitions. Molyneux's question is a thought experiment that asks whether a person who suddenly recovers from blindness can, by sight alone, tell the difference between the shape of a sphere and a cube when he has only ever distinguished these shapes by touch (Locke, 1689). In the temporal case, the question would be whether a congenitally blind person who suddenly regains sight can successfully differentiate between the temporal experience, e.g., the duration or temporal ordering, of two audiovisual events by sight alone, when he has only ever distinguished these temporal experiences by sound. E.g., can he differentiate or successfully recognise previously 'heard duration' when it is presented as a 'seen duration'?

There have been centuries of theoretical and empirical discussions of the answer to the spatial version of Molyneux's question, and the discussions as well as the results are equivocal at best (see Ferretti, 2020). But for the temporal version of Molyneux's question, the answer would seem to be a much more straightforward affirmative. The temporal version would not require the person to engage in any modality translation process. Comparing the duration would be a simple time-to-time comparison, even though the mechanisms underlying temporal information processing dissociate across modalities (see §3).

One natural explanation for why the temporal case is different may be due to the one-dimensional character of time. Unlike shape or spatial layout, which involve discriminating and comparing features across multiple dimensions, temporal comparisons across modalities may appear straightforward because temporal judgments just concern a single dimension. However, one-dimensionality alone does not suffice to explain comparability. The features mentioned above, such as intensity or distance, can be one-dimensional and yet not be effortlessly comparable across modalities. The ease with which we compare durations and temporal relations across modalities therefore cannot be explained solely by the dimensionality of time but still stands in need of explanation.

Beyond its phenomenological characteristics, comparability also plays a salient functional role. Consider the temporal ventriloquism effect. This effect is similar to the standard ventriloquism effect, where an auditory cue is illusorily perceived to originate from a more salient visual cue (Alais & Burr, 2004). The temporal ventriloquism effect shows how two temporally separated cues, a sound and a flash, are perceptually experienced as temporally bound at the time of the more reliable (in this case) auditory cue (Hartcher-O'Brien & Alais, 2011). It is not transparent to the perceiver which sensory signal is bound to which. The timing of each of the two modally distinct sensory signals is simply lost in the integration and perceived as a single unified perception. There are many other examples of how modally distinct features interact within our perceptual system to give rise to unified and reliable representations of the world (Alais & Carlile, 2005; Chen & Vroomen, 2013; Groen et al., 2013; Mégevand et al., 2013; van Wassenhove et al., 2007; Vroomen & Keetels, 2010; Zampini et al., 2005).

## 3 Fragmentation and Dissociation in the Mechanisms of Time Perception

While our perception of time appears unified in these three ways, the underlying mechanisms are far from unified. Viera (2020) identifies three dimensions along which these mechanisms dissociate: modality, temporal property type, and timescale. I expand on these dimensions by adding two further distinctions: temporal judgment task and temporal structure. Evidence for each of these is presented in turn to clarify the explanatory task of 'explicating the relation between the mechanisms underlying temporal binding and the content and phenomenology of our experiences of temporal binding.'

### 3.1 Modality

Experimental evidence shows how one can manipulate non-temporal aspects of an event to intervene on the perceived timing of sensory features of a specific modality type, without it intervening on the perceived timing of sensory features of a different modality type. Burr, Tozzi, and Morrone (2007) demonstrate how saccadic eye movements during visual perception decrease the perceived duration of a visual event occurring at the location and time of the saccade, while leaving the perceived duration of a co-occurring auditory event unaffected (Burr et al., 2007; Morrone et al., 2005). This shows how temporal discriminations of visual and auditory signals dissociate.

Fornaciai, Markouli, and Di Luca (2018) also show how a much-studied effect on duration estimation impairment is modality-specific. The effect in question denotes how the auditory duration estimation of a short target interval (100 ms), but not a long target interval (1000 ms), is impaired when a short distractor (250 ms) is placed between the target intervals (Burr et al., 2013; Karmarkar & Buonomano, 2007). If the distractor between onsets and offsets of the target intervals is increased (to 750 ms), this selective duration impairment effect disappears. Fornaciai et al. (2018) found a modality-specific version of this effect for a longer target interval (300 ms). They found that for the longer interval, the auditory duration impairment effect disappears but remains in place for visual duration judgments of similar intervals, making it possible to intervene in visual duration perception without intervening in auditory duration perception.

While this only shows that modalities can dissociate for duration judgments, there is also ample experimental evidence of how sensory features of a specific modality can modulate the perceived timing of sensory features of a different modality, resulting in cross-modal interactions. The ventriloquism effect described in Chapter 2 and above shows how certain modalities may interfere with each other to alter the temporal binding of those features. In this case, a visual cue is perceived as occurring synchronously with an auditory cue, regardless of whether that auditory cue occurs just before or just after the visual cue (Hartcher-O'Brien & Alais, 2011). Other evidence shows how looking at a person's lip movements has a large effect on the integration of visual and auditory features (McGurk & MacDonald, 1976). When attention is directed to visual lip movements that match the spoken sentence, compared to when the spoken sentence is non-matching, there is increased activity across brain areas, providing evidence for explicit multisensory integration activity for congruent vs incongruent auditory and visual cues (Senkowski et al., 2008; Vroomen et al., 2001). Vroomen & de Gelder (2004) similarly show how the flash-lag effect (where a brief flash is perceived as located behind a spatially co-located moving stimulus) is reduced when the flash is accompanied by a sound burst. In a different study, Groen et al. (2013) show how tactile sensory processing can influence early visual processing, causing the perceptual interpretation of when both these occur relative to each other to change. As such, neural processing in one modality can cause or even disrupt the perceived synchrony of sensory features from other modalities.

A theory of temporal binding should explain how these diverse ways in which sensory information from different modalities integrates and dissociates connect with our perceived unity of time. That is, one must account for how these modality dissociations occur and integrate such that they give rise to our unified perception of time, within which there is no such modal distinctiveness.

### 3.2 Temporal Property Types

The perception of distinct temporal property types, such as duration, sequences, synchrony, and rates of change, can be independently manipulated. Viera (2020) gives an example of this by referencing a study of a temporal oddball effect. In oddball paradigms, subjects are presented with a series of qualitatively identical stimuli, with a single oddball stimulus presented somewhere within the series (Tse et al., 2004). The oddball differs only in its non-temporal sensory features (size, colour, or shape). This change in non-temporal features causes subjects to perceive a change in the oddball's temporal properties; it is perceived to have a longer duration. Eagleman (2008b) uses this oddball effect to show how mechanisms encoding information about different temporal property types dissociate. In this study, the stimuli flicker at a specific rate, however, surprisingly, subjects' reports reveal that it is only the apparent duration of the stimulus that is subject to the oddball effect. The temporal judgment of the oddball's perceived flicker rate remains unaffected. Hence, the mechanisms underlying the perception of duration and rate of change dissociate (see Paton & Buonomano (2018) for a range of other dissociations relating to distinct temporal property types).

There is also a large body of evidence indicating that the perception of temporal order and synchrony relies on distinct mechanisms. Studies by Love et al. (2013) and Mégevand et al. (2013) show how synchrony judgments require a participant to judge whether a set of features occurs simultaneously, whereas temporal order judgments require a participant to judge which of the features came first, second, etc. In temporal order judgments, the point of subjective simultaneity (the point at which a subject is statistically most likely to judge two stimuli as simultaneous) is almost always audio-leading, while it is video-leading during synchrony judgments (Love et al., 2013). This kind of dissociation is widespread across many types and combinations of stimuli (Stevenson & Wallace, 2013; van Eijk et al., 2008; Vatakis et al., 2008)[20].

One would need to explain why one can intervene in the perception of one temporal property type without intervening in another. If time perception were unified due to the underlying mechanisms being unified, then intervening in the perception of one temporal property type should also affect the perception of a different temporal property type. One must explain how our

---

[20] Some doubt this dissociation due to experimental inconsistencies in stimulation, experimental set-ups, the distinct sets of biases in play, and the method of analysis (see García-Pérez & Alcalá-Quintana, 2015; Linares & Holcombe, 2014).

phenomenology of comparability and temporal binding unity is connected to, and can arise from, this dissociable set of discriminatory representational capacities underlying time perception.

### 3.3 Timescale

Viera (2020) points to discrete pharmacological and mechanical interventionist evidence by Rammsayer (1999) showing how temporal discrimination at different timescales can be selectively manipulated. Out of two drugs (haloperidol and midazolam) that both impair temporal discriminations around the 1-second timescale, only one (haloperidol) also impairs discriminations around the timescale of 50 ms. This fine-grained functional taxonomy of timescales in the mechanisms underlying time perception of sub-second and supra-second timing also gains support from psychophysical and neurological imaging data (Paton & Buonomano, 2018). Despite this, features are experienced within a single temporally bound moment that appears to be represented within a single timeline. There is nothing in our perception that makes a judgment about changes in the millisecond range different from changes in the second range, even though different mechanisms are employed for these discriminations. A theory of temporal binding should be able to account for how the disunification in the underlying mechanisms connects and gives rise to the unification of perceptual experience.

### 3.4 Temporal Judgment Task

A range of cases shows how our perception of temporal order and simultaneity is distorted by the kind of judgment or behavioural tasks we are engaged in. When we consider ourselves to be the cause of some effect, we tend to temporally bind these events (Haggard et al., 2002; Muth et al., 2022; Ruess et al., 2017; Schwarz et al., 2019; Tanaka et al., 2019). When features appear to be causally connected, we are more likely to temporally bind these events (Bechlivanidis et al., 2022; Bechlivanidis & Lagnado, 2009; Fornaciai & Di Luca, 2020; Kohlrausch et al., 2013). When we are continuously engaged in judging the temporal relation between two properties, we adapt to fixed delays between the features, causing the perceived delay to decrease or even disappear (Chen & Vroomen, 2013; Fujisaki et al., 2004; Roach et al., 2011; Rohde et al., 2014; Stetson et al., 2006; Vroomen et al., 2004). When we are looking at another person's lips and trying to interpret what they are saying, we are more likely to perceptually ignore lags between auditory and visual cues and temporally bind them into an audiovisual perception (on average, lags are not detected even when the soundtrack lags lip movement by ~140 ms or leads it by ~80 ms) (Dixon & Spitz, 1980; Hashimoto & Kumashiro, 2004; McGrath & Summerfield, 1985). How our attention is directed in different behavioural tasks is also important, as studies show that exogenous and endogenous attentional cuing of features greatly influences temporal binding (Chennu et al., 2009; Correa et al., 2006; Donohue et al., 2015; Holcombe & Cavanagh, 2008; McDonald et al., 2005; Senkowski et al., 2008; Talsma et al., 2010).

Accounting for the perceived unity of time would involve explaining how different task demands and behaviour can differentially influence time perception, while explaining why these effects remain introspectively inaccessible. Switching tasks, goals, or behaviour does not seem to result in a break in comparability or in which features are perceived as temporally integrated, even though task-switching interferes with our temporal discriminatory representational capacities.

### 3.5 Temporal Structure

Lastly, Nishida & Johnston (2002, 2010) and Arrighi, Alais, & Burr (2005) show how perceptions of synchrony and asynchrony depend on the temporal structure of the features in question. In Chapters 2 and 3, I explained how differences in temporal structure may matter for temporal discrimination. As I explained then, we discriminate first-order changes (e.g., sudden changes in an attribute like colour or position) at a much higher frequency and with more precision than second-order changes (e.g., motion reversals of a continuously moving object). See Chapter 2 for more details.

The fact that we process features with differing temporal structure matters because the discrimination of features with different temporal structures will be picked up with different levels of temporal precision. In cases where we have to make perceptual judgments about events that require us to integrate both kinds of changes, we may fail to accurately temporally bind the location of one feature undergoing a first-order change (colour change) with the location of another feature undergoing a second-order change (continuous motion) (Holcombe & Cavanagh, 2008; Nishida & Johnston, 2010).

Thus, how we temporally bind features depends on their temporal structure. It must be explained why discrepancies in tracking the timings of transients that differ in their temporal structure relate to our integrated and unified perception of time, such that we can explain how these differences in temporal structure affect what appears to us as a unified, integrated perception.

Taken together, all these different dissociations reveal a deep tension between our apparent experiences and some of the mechanisms underlying time perception. Along dimensions of modality, temporal property type, timescale, task demands, and temporal structure, the mechanisms that support temporal discrimination fragment in experimentally tractable ways. Yet none of these fractures are mirrored in experience. We do not perceive multiple modal times, task-relative timelines, or property-specific temporal formats. Instead, temporal features are experienced as bound to a single present, ordered within a single timeline, and effortlessly comparable across modalities and contexts. Any adequate theory must therefore explain how these fragmented and dissociable mechanisms give rise to a phenomenology that is stable, unified, and largely insensitive to the underlying fragmentation and dissociation. The challenge is not merely to show how temporal information is integrated, but to explain how the fragmentation, so evident at the mechanistic level, connects to our apparent experiences of the perceived unity of time.

## 4 Theories of Time Perception and the Perceived Unity of Time

Before proposing my own positive account, it will be instructive to first discuss how two kinds of theories of time perception might account for the explanatory challenge at hand. The two kinds of theories are the internal clock model and the mirroring approach. Although these two theories are very different, they share a common strategy when it comes to explaining the perceived unity, i.e., explaining the three perceptual capacities of temporal binding unity, timeline unity, and comparability. As argued by Viera (2019, 2020), both kinds of theories provide an account of how the posited representational vehicles responsible for processing sensory information attribute temporal content to the processed features. These theories describe these attributed temporal contents as a sort of common representational code that causes time to be experienced as unified.

### 4.1 Internal Clock Models and the Perceived Unity of Time

Internal clock models (also known as dedicated clock models (van Rijn et al., 2014)) were first developed by Treisman (1963) and Gibbon, Church, & Meck (1984)). According to these models, the contents of early sensory processing do not, by themselves, encode temporal properties such as duration or order. On this view, sensory systems primarily represent non-temporal features (e.g., pitch, colour, spatial location), and temporal content is not intrinsic to these early representations. Instead, these models take timing to be the result of a specific cognitive module, such that temporal properties are attributed at a later stage to whole perceptually represented events by a centralised timing mechanism, an internal clock. This internal clock is a cognitive system that monitors and tracks the temporal unfolding of the representations generated by different sensory processes. In this way, temporal experience is not locally represented within individual sensory modalities but is instead constructed by a supramodal mechanism that integrates information across modalities and processing stages.

The standard and most widely adopted internal clock framework is the Scalar Expectancy Theory (SET), developed by Gibbon, Church, & Meck (1984). The SET accounts for the apparent unity and systematicity of time perception by positing three core components: (i) a supramodal pacemaker–accumulator clock, (ii) a memory store, and (iii) a decision or comparator mechanism. Together, these components explain how organisms measure, store, and compare temporal intervals across different contexts and sensory modalities.

In the SET, the pacemaker (sometimes described as an oscillator) emits pulses at a roughly stable rate, though this rate is known to vary as a function of physiological and contextual factors such as attention, arousal, body temperature, pharmacological interventions, and heart rate (Behm & Carter, 2020; Gibbon et al., 1997; Schwarz et al., 2013; Van Volkinburg & Balsam, 2014). When timing is initiated, these pulses are gated into an accumulator, which counts the number of pulses generated during the relevant interval. The total number of accumulated pulses thus provides an internal, analogue magnitude representation of the duration of that interval. Longer intervals correspond to a greater number of accumulated pulses, while shorter intervals correspond to fewer pulses. The pulse count associated with an event is then stored in memory (in working memory for current intervals and reference memory for previously learned or standard durations) (Gibbon et al., 1984). Finally, the decision or comparator mechanism evaluates temporal judgments by comparing the accumulated pulse count for a current event with pulse counts retrieved from memory, enabling judgments of duration, simultaneity, or temporal order. This system can likewise represent order judgments of events by keeping track of intervals that separate the boundaries of events (Jones & Wearden, 2003). From this story about how temporal contents are attributed to events by a single coherent system, the internal clock can explain our perceived unity of time[21].

### *4.1.1 The Internal Clock and Temporal Binding Unity*

Internal clock theories can explain temporal binding by taking the clock mechanism to provide a common temporal ordering for all sensory information processed and encoded independently by a range of sensory processing mechanisms. There is only one system, the internal clock, that attributes temporal content to sensory features and, therefore, only one unified moment within

---

[21] Although scalar expectancy theory (SET) and related pacemaker–accumulator models originated in the late twentieth century, contemporary time-perception research continues to engage with internal clock frameworks in both behavioural and computational contexts (Allman et al., 2014; Allman & Meck, 2012; Sanabria, 2020; van Rijn et al., 2014; Wang & Wöllner, 2020). These contemporary internal clock approaches retain the core commitment that temporal information is represented in a common representational code by some dedicated cognitive module, which are the main tenets that are the target of this chapter.

which sensory features can be bound. This moment is directly dictated by the output of the comparator mechanism. Given that every sensory signal is integrated within a common temporal ordering by this mechanism, the theory can also easily define a very determinate moment as the time at which we experience different features and events to happen.

However, there is an issue with the fact that different sensory modalities process information at different speeds (see the problem of desynchronisation in Chapter 2, §3). To give just one example, we process auditory information much faster than visual information. Moreover, we have a much higher temporal resolution for picking up auditory signals than visual signals (Bullier, 2001). This is a problem because it might mean that two auditory features from two different events, E1 and E2, may be processed in the time it takes a visual feature of E1 to be processed. This might cause some issues in properly integrating the rapidly processed auditory feature of E1 with the more slowly processed visual feature of E1. What happens experientially is that the visual feature of E1 is integrated with the auditory feature of E2. However, it is not clear how a central clock can account for this and compensate for these differences, as the central clock does not attribute temporal content to specific features but to whole events.

One could account for such differences by positing local modality-specific clocks that accumulate and count pulses at the level of specific modalities (see Chen & Yeh, 2009). However, to account for how we experience temporal binding unity, one would still need some supramodal clock that translates the modality-specific temporal content into a unitary content attributed to whole events. This sort of solution merely kicks the can down the road, as there needs to be some mechanism that ensures that features of E1 are attributed the same temporal content despite different processing times of the features making up that event (at least when empirical evidence shows that this compensation for different processing times takes place, see Chapters 2 & 3).

A defender of internal clock models might reply that these dissociations do not undermine the view, since their effects can in principle be compensated for before temporal information is integrated by a central clock and within a single central timeline. On this picture, modality-specific distortions and delays merely affect where events are located on the master timeline, without threatening its unity. However, once the full range of dissociations discussed in §3 is taken seriously, this compensation strategy becomes increasingly problematic. To align temporal information across modalities, property types, timescales, task contexts, and temporal structures, the system would need to implement a large number of modality- and context-specific correction and translation mechanisms. These would have to continuously recalibrate temporal signals before they are integrated by the central clock. The resulting architecture would be highly complex, biologically implausible, and increasingly ad hoc.

Moreover, insofar as these compensatory processes must already encode detailed temporal relations locally in order to perform the necessary corrections, the explanatory role of the central clock becomes unclear. Either the clock merely aggregates outputs that have already been temporally structured, in which case it adds little explanatory value, or it must repeatedly re-encode temporal information that has already been encoded elsewhere, resulting in what Dennett has dubbed superfluous ‘double-transduction mechanisms’ (Dennett, 1988). That is, temporal contents are attributed to sensory contents at least twice (by the modality-specific clock and the supramodal clock). Introducing such double-transduction mechanisms makes keeping track of the timings of signals biologically overcomplicated and should strike one as implausible, especially when there are alternative theoretical solutions. In either case, the internal clock model fails to provide a satisfying account of how fragmented temporal processing gives rise to unified experience.

It is the many other examples of dissociations in the underlying mechanisms that make this issue pressing for the internal clock model. Consider how we perceive the temporal bound unity of a pattern that continuously changes colour and shape. Given the differential processing times of colour and shape features, we would need a clock that not only attributes temporal contents to specific modalities but to specific features in order to perceive whether these two features occur at the same time (see Chen & Yeh, 2009). Moreover, to account for the slightly distinct case of a dot that changes colour and motion direction, we need not just a system that attributes temporal contents to sensory features but also a system that attributes temporal contents to features that have different temporal structures (Nishida & Johnston, 2002). As mentioned in §3, the instant colour change is a first-order temporal change, while the motion-direction change is a second-order change. To have clocks that keep track of this difference in temporal structure, we would need to introduce ‘temporal-structure clocks’ that track transients produced by first-order changes and distinct clocks that track transients produced by second-order changes, and so forth.

The argument against positing new local clocks to account for temporal binding unity in light of new dissociation data is not that doing so would necessarily get the timings wrong, but merely that the abundant positing of different kinds of clocks becomes increasingly ad hoc. If we are to take the dissociation data seriously, then it seems that there is a wide range of dissociable mechanisms underpinning time perception, and positing a new clock for each of these mechanisms makes the internal clock view increasingly complicated and implausible. Either our brain needs to have pacemaker-accumulator systems linked to each mechanism across the dimensions mentioned in §3, or the clock is simply a proxy term for whatever kind of temporal processing is ongoing in the specific mechanism. The first option is improbable and an absurdly biologically taxing way of representing information, and it is not clear how it solves the issue of temporal binding unity. That is, it is unclear how the temporal information of these many different clocks could be integrated by a central system so that they are presented in a unified timeline. The second option is no longer an internal clock view, as it does not require any attribution of temporal contents by a secondary central clock system. Rather, if clocks are present at all levels, then the attribution of temporal contents becomes isomorphic to the time at which the sensory features are processed, i.e., the sensory features' temporal endpoints (see Chapter 2). In effect, such a 'solution' would collapse into a brain time theory, as it would mean that the temporal content of perceived events is determined by perceptual endpoints (the main tenet of the brain time theory). This theory comes with different issues (see below and Chapter 2).

*Timeline unity* seems to be similarly intuitive, and similarly problematic, to account for using the internal clock theory. This is because if there is just one common temporal ordering attributed by a central clock to whole events, then it is easy to account for how we perceive different events as occurring within a unified timeline. Events are neatly segmented from each other and related by being attributed to different times along the clock. These different, coherently represented elements could make up a consciously represented timeline that is maintained in working memory (Hsieh et al., 2011). But again, this explanation faces the same problem as the one for temporal binding unity. When we factor in all the kinds of dissociations that occur in the underlying processing, it is not plausible that a central clock could account for all the dissociations that occur within the attributions of temporal content to different modalities, features, temporal structures, etc., and as such it is also implausible as an explanation for how we represent a timeline as neatly unified. If we introduce multiple clocks at lower levels and a single central clock at the top, there is a problem of accounting for how we perceive events as neatly unified on a linear timeline, when all the temporal content attributed by these lower-level clocks would have been subject to differences in processing times, temporal resolution, and desynchronisation relative to when the features of events occur in the external world.

*Comparability* can also be given a seemingly straightforward explanation within the internal clock model. According to this model, all temporal contents of different pieces of sensory information are attributed by the same central clock. As such, all information about time will be of a unitary and comparable kind. However, as we have seen above, having a central clock runs into empirical issues, as it becomes unable to account for all the dissociations in temporal information processing. To remedy this, one could again embrace the view that we have a combination of many dissociable clocks. But this alone does not explain comparability, because now we have many dissociable temporal contents attributed by different clocks. To solve the issue of comparability, one still needs to posit a single internal master clock, such that temporal contents are attributed to events in a coherent and unified way by the perceptual system's final master clock (this again posits a superfluous double-transduction mechanism). With this combination of dissociable local clocks and a master clock in place, there will be no variation across modalities, timescales, property types, etc., for the temporal contents related to events. Thus, there will be no comparability issues. Everything that needs to be compared is already represented as a single unified temporal perception.

The issue is again the proliferation of these dissociable local clocks. Suppose one posits modality-specific clocks to account for the dissociation within the dimension of temporal information processing of modalities. To make this account plausible, it requires additional mechanisms, as the pulses of distinct clocks would pulse at different, faster and slower intervals, representing dissimilar durations. Viera (2020) mentions how this might be solved by having an in-built multiplication operation that normalises these pulses. Comparability is then assured by the fact that temporal information is represented by being converted into a single common code. Yet even more complicated solutions are needed to account for all the other dissociations. This strategy again begins to look unfeasible despite its initial appeal for the same reasons spelled out above. If we take all the dissociation data seriously, then we should not expect that one could give an internal clock-like story about the brain's temporal information processing. Although an internal clock model might be able to explain one or two of these dissociations by positing multiple local clocks, it is not able to account for all of them within a single internal clock model (see also Viera, 2020). This is because the baseline assumption for these kinds of internal clock models is that temporal contents are attributed to whole events by some centrally controlled process. One could present each of the dissociations cited in §3 as challenges that need to be overcome by the internal clock model. These many dissociations all point to the same flaw, which is that the internal clock model, as it stands, simply does not provide an empirically adequate description of temporal processing. The fragmented processing underlying time perception cannot be unified within a single temporal processing mechanism, nor can the proliferation of smaller clocks save the empirical adequacy of the model without it making the model increasingly ad hoc and biologically implausible.

## 4.2 Mirroring Approaches and the Perceived Unity of Time

In contrast to internal clock models, mirroring approaches argue that temporal contents are automatically attributed to experienced sensory features as we experience them. When a sensory feature is consciously perceived by being processed or represented by the relevant consciousness-making mechanisms, this feature is perceived as occurring the moment it is consciously available. In this way, mirroring approaches hold that the time of representing, and the time represented are isomorphic, i.e., they accept the thesis of temporal isomorphism (see Chapter 2). This also means that as long as there is synchrony in the time at which different experiences are represented, the representational content we subjectively experience will also appear as synchronous.

At first blush, mirroring accounts such as those provided by Dainton (2000), Foster (1991), Miller (1984), Mellor (1981), Phillips (2014b), and Rashbrook (2013a) provide an intuitive and elegant explanation of the different kinds of unity of time.

Let us first consider temporal binding unity. The act of representing occurs at specific moments in the world, and because the temporal content of our sensory representations mirrors the time of representing, we experience temporal binding unity as a result of synchrony in whatever representational mechanisms are doing the representing.

Timeline unity is also intuitively explained. That is, because the events of the world occur within a single timeline and our representational mechanisms pick up on these events, the act of mirroring explains why we experience these events as occurring within a single unified timeline. Our perception of a single timeline is inherited from the fact that our representations occur within the timeline of the world. By appealing to facts about working memory and how we consciously maintain multiple representations at the same time, the mirroring account can account for timeline unity.

Mirroring accounts might also be able to account for comparability. Because of the mirroring principle, the representing activity of different representational processes attributes temporal content to the individual representations of different sensory features simply through representing a feature. For this reason, any system that can make use of the sensory content encoded in one representational process (e.g., colour processing) should also be able to make use of the temporal content of those sensory features, since these temporal contents are determined by the time at which sensory processing finishes (see, Viera 2020, 645). There is, as such, no temporal comparison above that of spatial binding needed as this is undertaken automatically.

These three explanations are all satisfactory to the extent that something like the mirroring principle is true. The general criticism of mirroring approaches is, however, that they leave the mechanism underlying such a mirroring principle rather vague. As such, they might be taken to 'merely' propose the phenomenological claim that the time of representing and the time represented *seem* to stand in an isomorphic relation, and the claim that this would be an intuitive relation for time perception (Phillips, 2014b). This is then used to directly support the metaphysical claim that the two must be isomorphic. But this leaves us without any story about why this is so, what functional mechanisms are relevant for mirroring, and how one intervenes in them.

I take it that the brain time theory discussed in Chapter 2, could be understood as a way (perhaps the only way) to provide a plausible information-processing account of how this mirroring is accomplished. Instead of justifying this mirroring process based on our phenomenology, brain time theories justify this thesis by assessing the various ways our brain processes sensory information and the neural delays involved in this processing to make empirical claims about how neural processing times determine the time of representing. Recall that the brain time theory argues that the temporal contents represented in experience are wholly determined by the time at which a set of neural processes realise the conscious experience of some sensory feature. The time represented is determined by the time of representing. The brain time theory states that our perception of synchrony and asynchrony is determined by neural processing times.

If we take the brain time theory as a way of explicating the details about how the mirroring of mirroring accounts is accomplished, then these accounts do not have the same problem as the internal clock model concerning the dissociation data. According to this view, there is no issue in having a large array of dissociable timekeeping mechanisms. We simply perceive a sensory perception the moment the relevant neural mechanism reaches its perceptual endpoint. These mechanisms are free to dissociate in the many ways described in §3, and so the fragmentation of timekeeping mechanisms need not be an issue on the mirroring account. However, endorsing the brain time theory's account of the mirroring mechanism reinvites the problem of accounting for temporal binding unity. This is because if the time of representing and time represented are isomorphic, then experiencing the world in a temporally unified sequence of events requires that all the sensory features pertaining to the same event must finish processing at the same time. But as we argued in Chapter 2, this cannot be what happens (see Chapter 2 for details).

If the mirroring approach accepts the brain time theory's way of setting out the mirroring principle, then mirroring accounts can explain the fact that our mechanisms of fragmentation underlying time perception are fragmented. But this comes at the cost of being unable to account for our experiences of temporal binding unity as our experiences of unity are discordant with the order of sensory processing.

In conclusion, neither the mirroring approach nor the internal clock model will suffice, because they either explain the perceived unity of time by treating temporal information in terms of a central unifying mechanism (internal clock models), thus clashing with empirical dissociation data, or by appealing to temporal isomorphism (mirroring approaches), which, once fleshed out in any reasonable information-processing terms, leads to issues with accounting for our experiences of unity.

### 4.3 A Different Route to the Perceived Unity of Time

I want to rephrase how to deal with the explanatory challenge at hand. In accounting for the perceived unity of time, we should not look for unity in the respective representational vehicles. Instead, we should focus on how the contents encoded by these diverse timekeeping mechanisms might be integrated without the need to localise this integration within the bounds of any particular mechanism, or as the result of synchronous sensory processing.

I suggest that we can do this by looking to the hybrid theory of temporal binding presented in Chapter 3. According to this theory, sensory features might be perceived as temporally integrated without this integration taking place at any determinate moment or being performed by any distinct set of mechanisms that are directly responsible for encoding or representing temporal contents. Instead, our perception of the temporal structure of sensory features emerges from how attention is allocated to already processed and predictively processed sensory features. Accordingly, the representational contents representing features, objects, etc., need not have any common temporal contents attributed to them that can then be compared across dissociable mechanisms to underpin our perceived unity of time. The brain, in other words, actively constructs temporal order when doing so is relevant to behaviour and judgement, making the temporal organisation of processed sensory features an adaptive and flexible process rather than a fixed, automatic one.

By approaching the issue via the integration of the content of our perceptions rather than the integration between the mechanisms themselves, we can account for the perceived unity of time while retaining the fragmentation and dissociations of the underlying mechanisms. Importantly, this explanation entails a deflation of these perceived unities of time. This is because the hybrid model takes conscious time perception to be constrained by attentional capacities (a set of limited capacities, or at least a set of capacities that depend on limited resources). As a consequence, our perceived unity of time, which appears to be experienced as richly unified across our perceptual field, is more local and limited than we may tend to describe it.

## 5 The Hybrid Theory of Temporal Binding

In Chapter 3, I proposed the hybrid theory of temporal binding (from here on, 'the hybrid theory'). Its core claims can be understood in terms of two hypotheses about how retrodictive processing, predictive processing, and attention are integrated.

> ***Hypothesis 1:*** Our perceptual system uses retrodictive and predictive processing in mutually beneficial ways to engage in flexible temporally binding of sensory features.
>
> ***Hypothesis 2:*** Attentional modulation determines which sensory features become part of a conscious temporally bound perception by modulating said contents. This makes temporal binding adaptive to salient temporal patterns, goals, intentions, actions, and expectations.

First, this hybrid theory proposes that the brain temporally binds features by combining predictive processing and retrodictive processing. Predictive processes allow the perceptual system to generate future-directed expectations that help our perceptions keep up with the external world despite delays in neural processing. Retrodictive processes allow for our perceptual system to revise and reorganise perceptual content given sensory information processed later or changes in our attention. This helps smooth over potential errors in our predictions as well as redraft previous perceptual interpretations over time to fit our goals, intentions, etc.

On this view, prediction and retrodiction are not separate systems but dynamically integrated processes serving a common goal: To allow for fast, accurate, and flexible temporal organisation of sensory information.

Second, attention (whether driven endogenously by goals, intentions, causal inferences, etc., or exogenously by saliency) biases which predictive perceptual hypotheses become part of a consciously accessible perception, and which retrodictive revisions matter given changes in incoming sensory signals and one's goals, intentions, causal inferences, etc. Prediction and retrodiction may occur widely and unconsciously, but only attentionally modulated contents become part of conscious temporally bound perception.

This attentional component explains why temporal binding is capacity-limited, why some temporal illusions depend on attentional cues. Moreover, it explains why temporal binding can adapt to different contexts.

## 6 The Hybrid Theory and the Perceived Unity of Time

I propose that we can use the hybrid theory to provide a novel explanation of the explanatory challenge at hand. In §6.1, I set out how the hybrid theory explains the dissociations from §3. In §6.2, §6.3, and §6.4, I use the hybrid theory to account for temporal binding unity, timeline unity, and comparability.

### 6.1 The Fragmentary Structure of Timekeeping Mechanisms

In §3, I described five dimensions across which various mechanisms underlying time perception dissociate: modality, timescale, temporal property type, temporal judgment task, and temporal structure. This fragmented and dissociated structure fits with the hybrid theory, according to which temporal binding does not require there to be any unity within the representational vehicles.

Within the structure of the hybrid theory, there is no issue with having a fragmented system of asynchronous timekeeping mechanisms that employ their own representational and computational mechanisms at different times to fixate their respective content (fulfil their representational task in the perceptual system). According to the hybrid theory, all contents compete to grab the attention of the perceptual system and play a part in influencing what becomes a conscious perception.

It is not only the processing time and the kind of processing undertaken by the underlying mechanisms that determine the content of conscious time perceptions. It matters just as much which of the contents represented by these mechanisms make up a good perceptual model of the external temporal relations of sensory features toward which we are oriented, and the environmental context in which we are immersed. This context-dependent addition of the second hypothesis makes attention a central aspect of time perception and allows one to explain how the underlying mechanisms can dissociate while still giving rise to the experience of temporal binding unity, timeline unity, and comparability.

Attentional modulation of a subset of the processed sensory information gives rise to a conscious temporally bound perception, and there is nothing in this process that requires the contents of this temporally bound perception to be unified in the underlying mechanisms, processing times, or ways of encoding information. As such, the hybrid theory is consistent with the fact that our temporal processing is fragmented and dissociable. Below, I explain how we can use the hybrid theory to model how these fragmented and dissociable mechanisms underlying time perception give rise to experiences of temporal binding unity, timeline unity, and comparability.

### 6.2 Accounting for Temporal Binding Unity

When discussing temporal binding unity, it is important to remember that there is a difference between a conscious state that represents its contents as ‘now’ and a conscious mental state that occurs at now. While the internal clock model and the mirroring accounts take these two to be one and the same, they diverge in the hybrid theory. This is because our temporal perceptions are constructed and redrafted through predictive and retrodictive processing, and so the time of processing need not determine the time represented. A feature processed moments ago may be represented as ‘now’ along with a feature just processed.

As mentioned in §2.1, it is important to be able to consciously perceive some set of information as occurring ‘now’. According to the hybrid model, our temporally bound perceptions of such moments are goal- and action-oriented, and this means that choices are made about how to temporally integrate features in our environment. Due to attentional selection, perception is biased both bottom-up and top-down to consciously represent information that is useful for our perceptual purposes and to ignore information that is not relevant.

There is plenty of evidence for the fact that temporal binding is goal-driven in this way. As mentioned in §3, our judgments of which features seem to occur simultaneously differ depending on whether we are engaged in a temporal order judgment task or a synchrony judgment task (Love et al., 2013) and whether we see ourselves as the causal effect of some sequence of events (Haggard et al., 2002; Muth et al., 2022; Ruess et al., 2017, 2018; Schwarz et al., 2019; Stetson et al., 2006; Tanaka et al., 2019). Synchrony judgments also compensate for differences in processing times to perceive asynchronously processed features as synchronous, depending on the features and representational goal in question (Alais & Carlile, 2005; Chen & Vroomen, 2013; Mégevand et al., 2013; van Wassenhove et al., 2007; Zampini et al., 2005). Expectations and predictions influence our temporal order judgment so much that ingrained predictions can cause perceptual reversals in a sequence of events, even when the sequence of events is presented multiple times (Bechlivanidis et al., 2022; Hartcher-O’Brien & Alais, 2011). Lastly, temporal binding unity also seems to be predictively goal-directed to maintain perception-action synchrony (White, 2018).

In Chapter 3, I argued that all these cases require that we model temporal binding as a hybrid process that takes our attentional selection of top-down goals, intentions, and beliefs, as well as bottom-up environmental context, into account. Modelling temporal binding as such also explains why we can perceive two things occurring at different times in event time as occurring synchronously, i.e., the time of processing does not determine the time represented. Moreover, this theory also explains why people might experience two features changing synchronously in event time simply because of a difference in the features' temporal structure (§3). As mentioned in Chapter 2 §2.1, we perceive fast oscillations of co-occurring colour change (a 1st-order change) and motion direction change (a 2nd-order change) as changing asynchronously (Moutoussis & Zeki, 1997).

As already mentioned in Chapter 3, §4.3, this can be explained by the fact that the temporal structure of the features puts different demands on attentional selection. Holcombe & Cavanagh (2008) show how, under circumstances where attention is guided bottom-up by an external cue, our perceptual system can simultaneously access the representational content of both features (the colour and the motion direction), and then people see the features as changing together. This may be a consequence of the external cue eliminating the differences in temporal structure of the colour and motion-direction. By cueing them at a specific time, both features are picked up as a first-order change.

Importantly, this kind of model does not posit a time-keeping mechanism that attributes temporal contents to all the dissociable processed sensory features (as mirroring models do), nor does it posit some final temporal unifying representational mechanism that attributes temporal properties to the perceived multi-modal event (as the internal clock model does). Instead, the hybrid model provides a flexible way of modelling what contents the phenomenology of temporal binding unity will include. The moment that we perceive as 'now' would be a highly constructed and continuously redrafted moment that is a subset of sensory contents that might have been processed at very different times (some may be predicted, some stem from memory, and some are directly processed) and in very different ways (they are processed by a range of fragmented and dissociable mechanisms).

As a consequence, the representational content of our perceptions of this unified 'now' should be able to flexibly change depending on the goals we have and the temporal context we are in. For example, the stability and longevity of the perceived now should be sensitive to how frequently the external sensory features change and how our attention and goal-directed actions shift relative to those changes (this helps explain a range of studies, see Alais & Carlile 2005; Chen & Vroomen, 2013; Mégevand et al., 2013; van Wassenhove et al., 2007; Zampini et al., 2005). Instead of being solely determined by when different stimuli are processed in some representational vehicle, the content of our conscious perceptions of the ‘now’ also depends on a list of additional attention-grabbing aspects, including the temporal complexity of the stimuli (across the five dimensions), the saliency of different signals (see Hartcher-O’Brien & Alais, 2011), the temporal context (see Alais & Carlile, 2005), our expectations/predictions (see Bechlivanidis et al., 2022), and finally how the event develops relative to those predictions (as they may be retrospectively redrafted) (see Fujisaki et al., 2004; Herzog et al., 2020). Another consequence is that any coalition of sensory features represented in a temporally bound ‘now’ remains conscious until it is outcompeted by another coalition of contents relative to this range of attention-grabbing aspects (see Li et al., 2017).

Lastly, the parallel sensory processing undertaken by our brain massively outstrips our attentional capacities, meaning there is no chance that everything in our perceptual field can be attended to at once. This means that we cannot temporally bind all processed sensory features at once, suggesting that we cannot experience all processed sensory features in our perceptual field as a temporally bound and unified ‘now’. On this view, there would at any time be a lot of processed sensory contents that we do not consciously perceive as ‘now’. Some contents will not be attentionally modulated (due to not being salient, being irrelevant to goals, being in our peripheral vision, or being unexpected yet not attention-grabbing). Some situations may not require a conscious and perceptually mediated representation (e.g., a rehearsed and unmonitored sensory-motor task). While it may seem like we perceive everything in our perceptual field as unified within one synchronous ‘now’, this view predicts that this unified moment may be phenomenologically sparser than it might otherwise be, simply because we cannot apply attention to our entire perceptual field. I discuss this consequence extensively in Chapter 5.

### 6.3 Accounting for Timeline Unity

Beyond temporal binding unity, we also seem to experience our temporally bound perceptions as occurring within a unified timeline. Philosophers have often argued that, to be able to deliberate and act in accordance with a continuously changing world, we must be able to consciously experience some set of past, present, and potential future perceptions as temporally related (Chisholm, 1981; Dainton, 2000; Droege, 2009; Gallagher & Zahavi, 2001; Grush, 2005; Hoerl, 2013; Husserl, 1991; James, 1886; Le Poidevin, 2007; Lee, 2014; Mellor, 1981; Phillips, 2014; Reid, 1785). The difficulty is to connect this kind of phenomenology of a coherent extended timeline to the fragmented and dissociable structure of the mechanisms underlying time perception.

In discussing the temporal organisation of perception, Holcombe (2015, 820) proposes that ‘experience is frequently the product of organisational processes whose purpose is not to create an ordered timeline. Rather, simpler grouping and segmentation processes can be more important, with ordering sometimes only a byproduct or not occurring at all.’ A range of studies (Hartcher-O’Brien & Alais, 2011; Hommel & Akyürek, 2005; McGurk & MacDonald, 1976; Sinico, 1999) indicate that the brain prioritises representing salient groupings of sensory features over representing accurate temporal order. That is, the saliency of auditory and visual signals may spring from the fact that they fit together (for example, lip movements and phonological information), which in turn can cause us to integrate them into a fused perception at the cost of missing out on being able to perceptually represent the temporal order of the features in question.

In a similar vein, some have suggested that the temporal organisation of sensory information more broadly serves a common adaptive role, ‘To maximize the amount of useful information, at minimal costs, tailored to the observer’s current needs and circumstances’ (Akyürek, 2025, 1). This fits well with the hybrid theory, according to which temporal binding is a flexible and adaptive process that seeks to mitigate neural delays and desynchronised processing of sensory information relative to goal-, task-, and context-dependent factors.

Accordingly, it seems that, while it is very useful to group sensory information into segments that are significant for making judgments and performing actions, it is rarely useful to spend precious resources on maintaining the temporal position of every processed location of some sensory feature. Maintaining this information would be extremely taxing on attention and working memory and rarely plays an important role beyond the more general causal and semantic information that can be extracted from some event. As such, I agree with Holcombe (2015) that it is unlikely to be a built-in feature of perceptual processing to present us with an experience of a single unified timeline. Our experiences are more likely to be temporally grouped into salient and coherent temporal segments without a preference for representing a determinate timeline that represents the temporal relations of multiple features across time.

However, while Holcombe is on the right track with regard to how temporal perception is organised, experience does still furnish us with some (albeit a deflated) timeline representation. In one way, the ability to maintain representations across short intervals, involving the discrimination of successive stimuli within intervals of hundreds of milliseconds, intervals of almost a second, and even intervals of several seconds, is necessary for instigating consciously mediated actions

(Baddeley, 1993; Ninokura et al., 2004; Roberts et al., 2018; Tye-Murray et al., 2016)[22]. The working memory system is developed for the sake of attaining maximally integrated interactions between otherwise isolated and distributed unconscious representational processes of the brain (Dehaene et al., 2006; Dehaene & Naccache, 2001; Mashour et al., 2020). Moreover, working memory requires the simultaneous active maintenance of multiple items and some representation of their temporal relations (Hsieh et al., 2011). So, since most attended working memory activity also results in conscious experiences, and working memory requires active maintenance of multiple items and some representation of their temporal relations, our conscious experiences must actively represent the contents of a few moments in some sequential manner. This representation of the temporal order of a select few items can then be verbally reported or used to guide our actions in a goal-oriented way.

There are, of course, some strict temporal limitations to this sort of timeline experience. The temporal limitations/boundaries of how far into the future and past our timeline extends are limited by how beneficial the maintained representations are to the organism. As such, we should expect the length of a timeline to also depend on the accuracy of the maintained representations. Trapp et al. (2021) and Akyürek (2025) describe how we should expect the accuracy and degree of temporal integration of sensory features to drop as the complexity and temporal interval of the event in question increases. This puts a limit on how temporally extended such a timeline perception can be before it is no longer adaptive and no longer perceived.

As such, one consequence of the connection between working memory and the perceptual experiences of timeline unity is that timeline representations must be local and limited to maintaining some select portion of our processed contents. This is due to the capacity limitation of working memory (Dehaene et al., 2006). When this is combined with the hybrid view of how temporal binding is constrained by attention, the timeline unity we perceive between distinct temporally bound perceptions is constrained both by the capacities of working memory and attention. This also means that, if attention is local and shifts around all the time depending on top-down and bottom-up influences, our timeline representations would be similarly local and changing. The perceptual representation maintained in working memory will only give rise to brief and locally unified timelines to the extent that the contents are selectively attended to and fit well into a timeline representation. This fits with Holcombe's (2015) idea that experiences of a timeline

---

[22] Representation or rather constructions of longer timelines (tens of seconds and even minutes) will work in a very different manner that is much more constructive and less driven by perception. It is more likely that these are the results of our capacities of short-term episodic memory (Akyürek, 2025; Fontes et al., 2016) rather than being a direct result of perception.

are partly a by-product of the temporal organisation of perception. With large shifts in our attention, we lose the temporally bound perception of the features previously attended, leaving us unable to place them into a single determinate timeline together with the new temporally bound perception.

Moreover, if the hybrid theory is right, it is very unlikely (probably impossible) that we can have any perceptual experience of a timeline such that all the processed sensory features of a few events are maintained within a timeline across time and across our perceptual field. This is again because it is unlikely that we can simultaneously and continuously attend to all the sensory features in our visual, auditory, tactile, etc., perceptual fields, temporally bind these features, and keep track of how all the perceived features temporally relate across time. Although we do have explicit conscious working memory representations that put multiple sensory features in temporal relation to each other, I propose that this only happens for locally attended features at relatively small intervals (depending on how our attention shifts).

As such, my view is both in agreement and disagreement with Holcombe's view. I disagree with the idea that our organisation of experience is not oriented towards representing the temporal ordering of features and events. Though this is because I take it to be the job of working memory to maintain multiple features at once and utilise information about when they occurred to solve a range of behavioural tasks. This is perhaps compatible with Holcombe's claim that the temporal organisation is not aimed at temporal order per se, but aimed at organising contents in gestalts, if we understand gestalts as perceptual segments that are biased by what is attentionally relevant.

However, I do agree with Holcombe that the perceived temporal ordering is not necessarily determinate, such that our perception always allows us to consciously entertain what happens in what temporal order. I also agree that these deficits may be due to shifts in attention (Holcombe, Treisman, & Kanwisher, 2001). Shifts in our attention should break our perceptual experience of the relative temporal ordering. This is because when we shift attention we cannot compare the previous temporally ordered features with the temporal order of the new features in focus. Moreover, if we do not have time to attentionally sample the information required to capture the temporal order of the features (Nishida & Johnston, 2002), or if the cues given are not salient enough to give rise to efficient attentional sampling (Reeves & Sperling, 1986), then the timeline representation of such features would be either indeterminate or just inaccurate.

A helpful way of framing this deflated conception of timeline unity is by distinguishing between different kinds of timeline representation in terms of how richly timelines are encoded and how reliably these allow for conscious timeline representations. We can lay these out as four views, ranging from *rich* and *determinate* timeline representations to no timeline representations:

- ***Rich determinate timeline representation view:*** On this view, perceptual processing automatically yields a determinate ordering of events, such that sensory features are placed within a unified timeline representation. When we allocate attention to these sensory features, the richly encoded representation of their relative temporal positions can be recovered, allowing us to make a determinate and accurate judgment about their temporal order.
- ***Weak determinate timeline representation view:*** On this view, perceptual processing automatically yields a determinate ordering of events, but the recovered representation of this timeline may be inaccurate. When we allocate attention to the features, we will form a judgment about their temporal order, though it may be mistaken due to limitations of how we representationally recover this timeline.
- ***Weak indeterminate timeline representation view***: On this view, perceptual processing does not automatically yield a determinate ordering of events. When we allocate attention to sensory features, we may not always recover a timeline representation, and we may not always be able to make a determinate judgment about temporal order. This is because the temporal organisation of sensory features is not aimed at representing timelines per se, but at organising content into attentionally relevant gestalts, which may not always include a determinate order. On this view, while sensory features are experienced as belonging to a single timeline, we cannot always deliver judgments about temporal order, and when we can, these judgments may sometimes be indeterminate or mistaken. ‘Indeterminate’ here means that features appear temporally organised, yet closer scrutiny fails to yield a stable ordering.
- ***No timeline representation view***: On this view, there is no temporal ordering encoded at any level of processing, meaning that any timeline representation is in principle unavailable and, as such, timeline unity is entirely illusory.

The hybrid theory supports the *weak indeterminate timeline representation view*, predicting that experience presents events as occurring within a unified timeline while only locally and temporarily encoding sufficient information to support ordering judgments for attended contents. This is a deflation of our experience because, while our perceptions of temporally extended events may seem as if they are always perceptually placed within the same determinate timeline, it is, according to this view, not always the case. Even when events or features are represented within a determinate timeline, this will only be the case for a subset of attended sensory features or events that make up an attentionally salient ordered representation. This also means that there may not be any singular perceived timeline that remains coherent across our perceptions. Rather, there may be a bunch of locally coherent timelines that effortlessly and continuously flow in and out of consciousness, usurping each other as the environmental context or our top-down representational goals change. However, we may never notice any such inconsistencies because we cannot check if our timeline representations are right or not without shifting our attention back to items represented in the previously experienced local timeline (at which point it is already too late to check).

Put briefly, timeline unity is not a global, stable feature of consciousness, but a dynamic contingent product of attention-guided working memory. Moreover, it is a unity which, despite the superficial appearance of continuous unity, is constantly breaking and reforming due to attentional shifts and representational limits.

### 6.4 Accounting for Comparability

Lastly, we need to make sense of comparability within the hybrid theory. If the contents of our temporal experiences stem from a fragmented array of mechanisms, why does it seem so effortless to compare the temporal content of different sensory features? How can it be that we seem to consciously perceive time as a unitary representational kind, making any modal or other dissociations imperceptible?

According to the hybrid theory, all the various dissociations in the underlying mechanisms only matter insofar as they determine what kind of content is available to be attentionally modulated at some given moment. These fragmented mechanisms contribute to processing temporal information that can be made available through attentional modulation. At a functional level, predictive processes organise information across a hierarchy of mechanisms that contribute to processing temporal information to anticipate the likely temporal relation of sensory features, while retrodictive processing allows for the reorganisation of these expectations given new information, cues, context, tasks, goals, etc., that grab our attention, allowing the perceived temporal relations of sensory features to form an integrated best estimate.

Together, these processes aim to minimise the discrepancy between our representations and the external world through predictively modelling and retrodictively redrafting. In this way, the processes underlying temporal integration are oriented toward maintaining a single coherent temporal framing of the environment. This continuous construction and reconstruction of perceptual models relative to the world accounts for why conscious temporal representations coherently flow into each other over time.

But what causes the temporal content of these percepts to be experienced as a unitary kind rather than as a range of different kinds of temporal contents that match the many distinct underlying mechanisms? It might just be that we perceive the temporal information of these dissociable mechanisms as comparable because of the internal coherence upheld by the hybrid predictive/retrodictive system.

Temporal binding (as spelled out by the hybrid theory) can be conceived as a distributed consumer system that presents attentionally modulated representational contents together. This does not mean that all processed sensory features become temporally bound (due to limits of attention), nor does it mean that the accessed contents are unified within some neural mechanism. The contents of the mechanisms underlying time perception remain fragmented and dissociable, and there is no system that attributes common temporal contents to sensory features before or as they are made conscious. The temporal binding process simply attentionally modulates those contents that allow for a probable and coherent conscious perceptual representation of the external world, and in doing so, we are presented with a range of contents that appear to represent the world as unified. The represented coalition of content seems to share a unitary kind of temporal content. However, their internal coherence might just make these contents appear to share some common temporal property.

If this is true, comparability is not attained in the same way as it is according to the internal clock models or other mirroring accounts. On this account, comparability does not depend on treating time as a common code that is attributed to every piece of representational information (the internal clock model) or directly importing temporal content from the time at which sensory features have been processed (the mirroring model). The different representational mechanisms encode their content in fragmented ways, only to be selectively modulated by attention in ways that form coherent conscious perceptual experiences.

Despite remaining fragmented, the representational contents appear temporally unified in the way they make up a locally coherent temporal representation of the world (with all the caveats of how temporally bound perceptions are flexible, attention-dependent, capacity-limited, and possibly incoherent across time depending on how one's environmental context changes). Attentional modulation of how conscious perceptual contents are predictively drafted and retrodictively redrafted ensures that our temporal perceptions remain locally comparable in this way and prevents information that conflicts with this coherent representation from being attended to and, thus, consciously experienced. This entails a deflation of our experience of comparability to the degree that comparability is local and limited in the ways that attention is local and limited. Also, this deflation does not mean our time-sensitive behaviours are limited to being guided by these attentionally constrained local temporal perceptions. Content that is processed and unconscious can still be used to unconsciously guide other time-sensitive behaviours.

In the following studies, we can see how attention can cause the content of otherwise isolated mechanisms underpinning time perception to influence each other as they become part of a unified temporally bound percept, without providing any introspective access to how these mechanisms influence each other. McDonald et al. (2005) show how exogenous attention allocated to a sound causes the participant to temporally bind visual features differently than they would have been had the sound not been present. In another study by Stetson et al. (2006), deploying endogenous attention to your own action's effect can cause a reversal of your temporal order perception. The temporal ventriloquism effect in Hartcher-O'Brien and Alais (2011) and Chen & Vroomen (2013) described above, similarly shows attention unifies the temporal experience of sensory stimuli from different modalities into what is perceived as a single coherent event occurring at one moment.

What about the case by Tse et al. (2004) (presented in §3) that Viera (2020) takes to count against the internal clock model? Participants presented with a regular sequence of stimuli (dots of light) perceived an oddball stimulus occurring in the sequence (that differed only in some non-temporal way) as having a significantly longer duration than the standard stimuli. This was a problem for the internal clock, as the apparent duration increase must be caused by a distortion of the pulse rate of the internal clock. But if this is the case, then this should affect the duration measurement of other co-occurring features. For example, if the stimulus and the oddballs were flickering at a constant rate, then the flickering rate of the oddball should be equally distorted, but this is not what happens (Eagleman, 2008b).

On the hybrid view, accounting for this is not an issue. It takes the attentional access or modulation of content to be the determining factor of our time perception, rather than the attribution of temporal contents by some internal clock. If attention separately accesses the content of the duration and the content of flickers, then adaptations to each of these features can freely dissociate, because this happens at the level of the mechanism. As such, the oddball might affect one timekeeping mechanism without affecting another and thus affect mechanisms locally without affecting all temporal content that is consciously perceived. The effect of the oddball may affect distinct mechanisms (see Johnston et al., 2006), and thus affect the kinds of contents that can be accessed about the duration, without affecting the contents of the flicker rate. With the hybrid theory, we can account for local distortions of temporal perceptual capacities without this giving rise to any noticeable difference in our subjective experience of time as being unified.

## 7 Conclusion

Although it may currently be difficult to give a more detailed account than the one above, I think that it is possible to meet the explanatory challenge of connecting our perceived unity of time to the fragmented processing underlying time perception. The hybrid theory of temporal binding provides a promising framework for understanding how we experience time as unified, despite the widely fragmented and dissociable mechanisms that underlie temporal perception. By conceptualising temporal unity in terms of attentional modulation rather than a common neural code, this approach accounts for the complex interplay between sensory and cognitive processes that shape our experience of time. While we can use this to account for our temporal experiences, the theory also entails several deflationary consequences.

We can explain temporal binding unity as the result of a goal-directed process, where perceptual selection determines what is experienced as ‘now,’ rather than being dictated by strict temporal isomorphism between processing and representation. This allows for constructed and redrafted perceptions of simultaneity, with attention shaping what features are bound together depending on context, salience, and predictive modelling. The perceived 'now' is thus a dynamic subset of sensory contents that flexibly adapts to shifting goals and environmental demands. The phenomenological temporal binding of the ‘now’ that we experience is, according to this model, less precise, more flexible, and sparser than we tend to describe it. Its contents and temporal breadth are determined by attentional limitations and constraints, both bottom-up and top-down.

Extending this theory to timeline unity, I have argued that while we may consciously experience a structured timeline of past, present, and future events, the hybrid theory argues that this timeline representation is both local rather than global and indeterminate and contingent rather than determinate and necessary. This is due to the fact that the temporal organisation of perception is dependent on how our attention is directed as well as the resource constraints of working memory and attention. At best we maintain a sequence of short, coherent temporal segments of a subset of processed sensory features. While representations of these segments can include information about temporal order, this is not a necessary feature of our representation of such segments. The claim here is that our temporal organisation does not select contents specifically to represent a precise temporal ordering, but rather to represent some coherent temporal grouping of sensory features in ways that are ecologically relevant for making judgments and performing actions. Another consequence of this view is that these represented temporal segments are continuously overwritten as our attention shifts relative to representational goals and the context at hand. This view accommodates both the apparent continuity of our timeline perception and its fragmentary underpinnings, suggesting that our experience of a unified temporal order is contingent, not intrinsic, to perceptual processing.

Finally, comparability can be explained within the hybrid theory as arising from the coherence of attentionally modulated representational contents. Despite the fragmented and dissociable nature of the underlying mechanisms, the predictive and retrodictive processes that shape conscious perception ensure that temporal contents appear internally consistent and comparable. Rather than imposing a singular temporal structure onto experience, the hybrid model explains how attention-driven selection and predictive modelling create an experience of comparability. Yet comparability is deflated to the extent that the comparable properties are limited to those we can simultaneously attend to.

If the hybrid theory of temporal binding is on the right track, then we must more carefully investigate how attention constrains our perceptual experiences of temporal unity. That is, we must investigate whether our experiences seem phenomenologically richer and more globally temporally unified than they could be if they were constrained by attention. In the next chapter, I set out a novel theoretical framework that structures how we could empirically test this proposed relation between attention and the phenomenology of our temporally bound perceptions.

# Chapter 5

## How Phenomenologically Rich Is Temporal Binding?

**View all animations referred to in this chapter here**[23]

### 1 Introduction

In Chapters 3 and 4, I have argued that if the hybrid theory of temporal binding is on the right track, we should expect attention to constrain both the content and phenomenology of our perceptual experiences of temporal binding. In this Chapter, I discuss in detail how we can empirically investigate these claims of the hybrid theory, by providing the best possible framework for empirically understanding how attention, the content of our experiences of temporal binding, and the phenomenology of temporal binding relate.

When we consult our phenomenology, everything in our conscious perceptual experience appears, at first blush, as temporally unified. All sensory features in our perceptual field are experienced as temporally bound and perceptually anchored in one synchronous moment in a unified way. This experience, which we can refer to as the *apparent phenomenology of global temporal binding*, suggests that our phenomenology of temporal binding extends across our entire perceptual field. The word *apparent* is meant to express agnosticism about whether this global phenomenology of temporal binding is a veridical description of our conscious experiences or a misdescription of our conscious experiences.

Multiple philosophers have noted relevantly similar kinds of phenomenological unity, though under different guises, such as synchronic unity, co-consciousness, and phenomenal unity (Bayne, 2010; Bayne & Chalmers, 2003; Prinz, 2012; Rashbrook, 2013b). Though these authors do

---

[23] https://rpthesis.wordpress.com/

not necessarily agree on how global this experience of unity is[24], they claim that when we have conscious perceptual experiences, all parts of these experiences are phenomenally co-experienced i.e., as consciously perceptually unified.

Understood in these terms, the apparent phenomenology of global temporal binding suggests that our phenomenology presents all sensory features and parts of our perceptual field as being co-experienced as temporally bound. Another way to describe the apparent phenomenology of global temporal binding is the absence of temporal binding. That is, there is a comprehensiveness to temporal binding such that no parts of our perceptual field, or any specific sensory features in our perceptual field, that do not seem temporally bound with all other parts and features.

This apparent phenomenology of global temporal binding is in tension with the hybrid theory of temporal binding in the following way. While we may tend to report that we have experience as of global temporal binding, the hybrid theory proposes that both our ability to perceptually represent features as temporally bound and our phenomenology of temporal binding are limited by attentional capacities (see Chapters 3 and 4)[25]. If both the content and phenomenology of our experiences of temporal binding are attention-dependent in this way, then our phenomenology of temporal binding could not be as of global temporal binding.

That is, features in our perceptual field could not be phenomenologically experienced as temporally bound if those features are not attended to. They could not even be experienced as such in an illusory way because an illusion of temporal binding requires us to represent the features as if they are temporally bound when they are not. According to the hybrid theory, even such illusory experiences are attention dependent.

As such, the theory states that any discrepancies between our phenomenology of temporal binding and representational capacities of temporal binding would be due to some misdescription of our phenomenology.

---

[24] Both Bayne and Dainton are for example explicit in arguing that the content that is perceptually experienced as unified can differ depending on various factors such as attention, working memory, etc. meaning that our experience of phenomenal unity can be more or less rich.

[25] As in chapters 3 and 4 I use the term attentional modulation as a way to talk about how information processing and the deployment of cognitive resources are altered selectively to change the kind of access that different mechanisms have to certain contents and incoming stimuli. Broadly put, I take attention to subserve the role of 'mentally selecting content for behavioural purposes' (Wu, 2024), which according to William James 'implies withdrawal from some things in order to deal effectively with others' (James, 1890, 404).

As I discuss in this chapter, this may take place due to a series of (systematically biased) subjective inflations, that cause us to describe our phenomenology of temporal binding as richer than it is. If this is the case, then we should deflate our phenomenology of temporal binding, holding that our temporal experiences are less rich than we tend to describe them as being. We can call this view *attention-dependent phenomenology of temporal binding*.

On the other hand, the hybrid theory may be wrong. It could be that we actually do have a phenomenology of sensory features as being temporally bound across our entire perceptual field (veridically or illusorily). This could be the case even if we cannot make (accurate or false) determinate judgements about how all these features are temporally related. In this case, our phenomenology of temporal binding overflows our ability to perceptually represent features as temporally bound.

This could be because the latter (the accessible representational content of our perceptions of temporal binding) depends on limited attentional capacities, whereas the former (our phenomenology of temporal binding) does not. On this view, our phenomenology of temporal binding is pre-attentive, even if our perceptual representation of temporal binding and related discriminatory capacities are not. We can call this view *pre-attentive phenomenology of temporal binding*.

Because of an absence of empirical studies on the phenomenology of temporal binding (as opposed to the representational content of temporal binding), it is unclear whether the hybrid theory is right about how our phenomenology of temporal binding depends on attention. However, exploring this relation is important to settle facts about which theory of temporal binding is on the right track.

According to the brain time theory explored in Chapter 2, temporal binding depends on the time at which sensory features finish processing and reach the processes responsible for conscious experiences.

If conscious perception itself does not depend on attention, but only on early recurrent sensory processing (Block, 2019; Lamme, 2007), then, according to the brain time theory, we can have experiences of temporal binding that do not depend on attention. As such, the brain time theory fits well with the pre-attentive phenomenology of temporal binding view, while the hybrid theory is only compatible with the attention-dependent phenomenology of temporal binding view.

In this chapter, I provide a first important step in how we can empirically investigate this relation between our experience of temporal binding and attention, by exploring the following three questions:

> ***Q1:*** Does the apparent richness of people's phenomenology of temporal binding depend on attention?
>
> ***Q2:*** Is there a discrepancy between how both: (a) the apparent richness of our phenomenology of temporal binding and (b) our capacity to engage in temporal binding tasks[26], depend on attention?
>
> ***Q3:*** If there is a discrepancy between how (a) and (b) depend on attention, then what explains this discrepancy?

Question 1 concerns what people's experiences of temporal binding are like. This question is interesting in and of itself, as there are no empirical studies of people's characterisation of their phenomenology of temporal binding (how rich vs. sparse they are) nor any research on how the phenomenological richness of such experiences is influenced by interventions in attention. Existing empirical studies on temporal binding monitor people's temporal discriminatory capacities concerning a specific set of stimuli and have very limited report paradigms. There is little or no emphasis in these studies on how people's experience of temporal binding relates to their performance on the given tasks.

To properly assess the richness of people's experience of temporal binding, studies need to allow for more open-ended report paradigms that can capture a much broader array of people's potential, or at least purported, phenomenology. In examining this question, I look to existing expansive report paradigms to propose a way to develop new studies that inquire carefully about people's phenomenology of temporal binding without simply taking performance on temporal binding tasks as a proxy for phenomenology.

Question 2 is comparative and concerns analysing trends in how the phenomenology of temporal binding and our discriminatory temporal binding capacities, respectively, depend on attention. This question is crucial in assessing theories of temporal binding. This is because if there

---

[26] By engagement in temporal binding tasks, I do not necessarily mean that our perceptions of temporal binding are accurate (get the relative event time relations of features right). I just mean that we perceptually represent features as temporally bound (veridically or illusorily) in such a way that this gives rise to the various actions, deliberations, reports, judgments, and other functions, that flow from (veridically or illusorily) temporally binding certain features and which do not flow from a failure to temporally bind them.

are any discrepancies in the trends in how our discriminatory capacities and our phenomenology of temporal binding depend on attention, then this puts empirical pressure on the hybrid theory.

Lastly, Question 3 concerns how we might interpret potential discrepancies in how our discriminatory capacities and our phenomenology of temporal binding depend on attention. If empirical research discovers such discrepancies, it could either be that such discrepancies are the result of the existence of genuine phenomenal overflow of our phenomenology over our discriminatory capacities (meaning that the hybrid theory is false). Or it could be that these discrepancies can be explained away as a result of various kinds of subjective inflations (supporting the hybrid theory). I explore a range of ways in which we tend to subjectively inflate our perceptual phenomenology to provide explanations and examples of how we might tend to inflate our phenomenology of temporal binding. I then suggest how we can put these subjective inflations to the test.

In §2.1, I set out a distinction and a methodological issue from the debate about theories of consciousness to lay out how one may empirically investigate the way perceptual content and perceptual phenomenology relate to attention. In §2.2, I use this distinction to set up the attention-dependent phenomenology of temporal binding view and the pre-attentive phenomenology of temporal binding view. In §§3.1, 3.2, and 3.3, I explore the three questions and related hypotheses in detail and develop a series of animations and psychophysical tests that may be used to get new data that help confirm/disconfirm such hypotheses. I conclude in §4.

## 2 How Perceptual Content, Perceptual Phenomenology, and Attention Relate

In this section, I introduce two broadly distinct positions on what conscious perception is, in order to clarify two competing views on how attention relates to perceptual representational content and perceptual phenomenology. I do this to provide an overview of the common methodological issues that stand in the way of making progress on questions concerning how attention may or may not constrain phenomenology, and to set realistic expectations for what kind of progress can be made through future empirical investigations. This sets up the conceptual framework for how we should go about making progress on the three questions set out above.

### 2.1 Two Distinctions on How Attention Relates to Conscious Perception

Theories of consciousness often pay particular attention to how conscious perception relates to attention, and are frequently divided according to whether they allow for dissociations between these to go both ways or only one way (Pitts et al., 2018)[27].

Double dissociation views claim, first, that subjects can deploy visual attention without this resulting in conscious visual perception of the target, and, second, that subjects can have conscious visual perceptual experiences of some features without deploying visual attention (Block, 2011; Koch & Tsuchiya, 2007; Lamme, 2003). These authors argue that there is evidence which suggests that people can have conscious awareness of the gist of a scene in the absence of attention. These gist perceptions include experiences of the summary statistics of a scene such as the colour changes within it (Bronfman et al. 2014) or general orientation of objects in a scene (Alvarez & Oliva 2008; Cha & Chong, 2018), and take place in the near absence of attention, or perhaps even independently of attention. Moreover, the case of blindsight mentioned in Chapter 3 shows how people can deploy attention to luminance and general motion of luminance patterns, increasing perceptual discriminability, without it resulting in any conscious awareness of the attended features (Kentridge, 2004).

Single dissociation views, by contrast, claim that although subjects can deploy visual attention without it resulting in a conscious visual perception (e.g., blindsight), they cannot have any conscious visual perceptions without deploying visual attention (Cohen et al., 2012, 2016; Naccache, 2018). They thus deny that studies on, for example, gist perceptions are evidence for there being conscious awareness in the absence of attention. In response, they claim that attention actually does constrain gist perceptions and that when attention is fully consumed by some central object or task, we would not have any awareness of the gist of a scene.

The disagreements between single and double dissociation views are as such partly based on empirical evidence of how attention and conscious awareness dissociate and partly based on phenomenological claims about how rich experiences seem relative to how constrained attention is. I assess the evidence on gist perceptions in §3.1.1 below to discuss how attention and the richness of our phenomenology relate.

At this level of description, the disagreement concerns whether attention is merely one of many factors that influences conscious perception, or whether it is a necessary condition for

---

[27] A minority have also held that attention just is the main mechanism of consciousness, with no dissociation between the two (Posner, 1994). However, for the present purposes we can ignore such views.

conscious perception. To make this distinction clearer, however, we must examine more carefully how these views map onto what different theorists mean by 'conscious.'

Many defenders of the single dissociation view take it to be necessary that if a brain state is a conscious perceptual state, then that state is also cognitively accessed. Let us call this hypothesis the *access-consciousness hypothesis* (there are many who endorse this broad definition of the hypothesis, though they differ considerably on the details, see Baars, 2002; Brigard & Prinz, 2010; Dehaene, 2014; Dennett, 1991; Lau & Rosenthal, 2011; Naccache, 2018, p. 201; O'Regan & Noë, 2001; Prinz, 2010, 2010).

On this view, a mental state counts as conscious only if it is accessible in a way that makes it usable in (i) deliberation, verbal reporting, consciously guided action, (ii) non-verbal reporting (e.g., button-pushing) guided by perceptually online representations, or (iii) no-report paradigms[28] (Naccache, 2018). Conscious perceptual states are therefore states whose content is representationally available. According to most who endorse this hypothesis, a minimum requirement for cognitive access is attention, meaning that attention is necessary (but not sufficient[29]) for conscious perception. As a result phenomenology and representational content are tightly linked. Both depend on attention, and neither can occur without it. There is, on this view, no principled distinction between what is phenomenally present in conscious experience and what is representationally available in our conscious cognition.

Advocates of the double dissociation view often reject this conception of conscious perception. They deny that access-consciousness exhausts conscious experience, and argue that cognitive access, and thus attention, is not necessary for conscious perceptual phenomenology. According to this alternative conception, there exist conscious perceptual states that are not

---

[28] These are cases where strong causal associations between neural signatures and reportability are used to gather evidence of conscious evidence from non-reporting subjects

[29] We can attend to things, e.g., visual events, without these being cognitively accessible. Beyond the case of blindsight, there are many examples of what Dehaene et al., (2006) call non-conscious accessible states and non-conscious subliminal inaccessible states which are non-accessed states that show evidence of relying on some degree of attention (Dehaene et al., 2006, 207). According to global workspace theories for example, non-conscious accessible states are a type of neural processing that potentially carries the right kind of bottom-up activation for becoming accessed states but must meet additional conditions e.g., stronger activation of higher associative cortices, particularly parietal, prefrontal and anterior cingulate areas—this includes attention (Dehaene et al., 2006). These accessible states might never be accessed, if they disappear from some cognitive buffer before they pass some threshold for becoming accessed, yet they can be attended and clearly contribute to priming effects (Cohen et al., 2012; Dehaene & Changeux, 2011; Naccache et al., 2002). There are also inaccessible subliminal states that rely on subliminal processing, which refers to a non-conscious type of neural processing whose informational content will remain consciously inaccessible because it does not carry the right kind of bottom-up activation being accessed, which nevertheless can be attended to such that attention plays a role in forced-choice paradigms (Dehaene, 2014, p. 201)

accessed, not directly available for a range of discrimination tasks and verbal reports (though views differ on how they understand this view, see §2.1.1 below). Let us call this position *the merely phenomenal consciousness hypothesis* (Block, 1995, 2007; Chalmers, 1996, 2010; Lamme, 2006; Lamme, 2003; Robinson, 2004). On this view, there are what we can call *merely phenomenal conscious perceptions*. These are perceptual states for which there is something it is like to be in them, even though their content is not cognitively accessed. Those who endorse the merely phenomenal consciousness hypothesis appeal to empirical studies indicating the existence of phenomenal conscious perceptual states (e.g., gist perceptions) that rely on early attention-independent activity in the visual cortex (Aru & Bachmann, 2013; Iwasaki, 1993; Koch & Tsuchiya, 2007; Scholte et al., 2008; Tsuchiya & Koch, 2016).

On the merely phenomenal consciousness hypothesis, while attention may be required for many kinds of perceptual representation and cognitive access, it is not necessary for perceptual phenomenology itself. As a result, subjects may have experiences whose phenomenal character extends beyond what they seem to be able to attend to, report, and perceptually discriminate in a range of psychophysical discrimination tasks.

A central motivation for the merely phenomenal consciousness hypothesis is the claim that appeals to access-consciousness alone cannot account for the apparent richness of perceptual experience. Access-consciousness depends on mechanisms such as attention and working memory, which have been argued to be limited to only a few items at a time (Cowan, 2010; Franconeri et al., 2007; Pylyshyn & Storm, 1988; Yantis, 1992). I discuss this more in §3.1. Many find it implausible that such limited mechanisms could underwrite the richness and detail of our perceptual experiences. However, adherents of the access-consciousness hypothesis are sceptical of this conclusion. They argue that there is no evidence for merely phenomenal conscious states that cannot be explained in terms of access-conscious states (Cohen & Dennett, 2011; Dennett, 1991; Frankish, 2016; Naccache, 2018).

### *2.1.1 A Methodological Issue of Testing the Relation Between Attention and Phenomenology*

A major obstacle in the debate about how attention and our phenomenology relate is methodological. Interpretations of psychological experiments (e.g., binocular rivalry, inattentional blindness, change blindness, subliminal perception, blindsight) that track people's performance and experiential reports often underdetermine choosing one theoretical explanation over the other (Block, 2007, 2019). Even when one factors in concurrent data from no-report paradigms, such as brain scans from binocular rivalry studies (Frässle et al., 2014; Odegaard et al., 2017; Safavi et al., 2014), people still differ in whether or not these provide evidence of one hypothesis over the other.

More generally, it is unclear how to design empirical tests that can decisively adjudicate between these positions. Both direct and indirect strategies rely, in one way or another, on behavioural or verbal reports, which themselves may presuppose cognitive access. As a result, evidence that is taken to support phenomenal overflow by one camp is often rejected by the other, and disagreements frequently reflect deep theoretical commitments rather than clear empirical divergences (Block, 2007, 2019; Cohen & Dennett, 2011; Lamme, 2003; Naccache, 2018).

To this end, it is important to distinguish between an extreme and a moderate version of the merely phenomenal consciousness hypothesis. This is because there are good philosophical reasons for not considering what I call the extreme version of this.

On the extreme version, merely phenomenal conscious perceptions are in principle non-functional, non-representational, and causally inert. They play no role in reasoning, report, action, memory consolidation, or any other downstream process. As such, they are not merely difficult to detect empirically, they are empirically intractable by definition. No behavioural paradigm, no neural signature, nor any report or no-report paradigm could ever provide evidence for their existence, since by hypothesis they make no causal difference to anything that could be measured. This extreme position is closely related to traditional forms of epiphenomenalism about phenomenal consciousness (e.g., Jackson, 1982; Chalmers, 1996).

There are, however, strong reasons to resist this extreme formulation and few people take this extreme view. The strongest reason for those who think that research on consciousness and conscious perception requires empirical evidence from psychology, neuroscience, and cognitive science is probably that taking this extreme view automatically lands the debate in an empirical deadlock. There would, by definition, be no prospects for new empirical or methodological insights in the brain sciences that could inform our view on the nature of this phenomenology.

Two other reasons for rejecting the view come from long-standing arguments in philosophy of mind. First, the extreme view cannot be straightforwardly motivated by appeal to conscious phenomenal experiences themselves. If merely phenomenal experiences are strictly causally inert, then they cannot explain our beliefs about them. This is a familiar challenge to epiphenomenalism in general. If phenomenal states play no causal role, then they cannot cause our reports, our introspective judgements, or even our theoretical convictions about their existence.

Second, because no appeals to empirical studies or to phenomenology can motivate this view, one's commitment would have to be grounded in some theoretical intuitions. One might have the following intuition. If we gradually 'turn off' the mechanisms of cognitive access (attention, working memory, reportability), it seems implausible that conscious experience would simply disappear when we remove the last of these functions. Conscious experiences, on this line of thought, should not switch off merely because the last function associated with such conscious experiences has been removed. Conscious experiences could as such be free-floating from their otherwise downstream effects. But this is nothing more than an intuition about the metaphysical independence of phenomenology from function. It does not by itself provide or allow for any external motivation.

While such a view is possible to hold, I see little motivation for holding it. Moreover, it is entirely uninteresting relative to a debate concerned about how we may provide the best possible empirical framework for understanding how attention relates to our phenomenology of temporal binding.

For these reasons, I set aside this extreme, empirically intractable, version of the merely phenomenal consciousness hypothesis. Instead, I focus on a more moderate and empirically tractable version. According to this weaker view, attention and cognitive access are not necessary for consciousness because phenomenal consciousness exceeds access-consciousness. This, however, does not mean that phenomenal consciousness is causally inert. It causes us to have rich experiences that we can report something vague about, yet not in ways that translate to or match functional downstream effects. Moreover, the claim is that the theoretical considerations offered by those who adhere to the access-consciousness hypothesis cannot account for these experiences. It is this moderate merely phenomenal consciousness hypothesis that is both of philosophical and empirical interest, because it gives rise to a debate about which mechanism we must appeal to in order to account for the full extent of our conscious experiences.

#### *2.1.2 Collecting and Respecting Phenomenological Data in Theoretical Explanations*

In response to the empirical underdetermination of the choice between the moderate merely phenomenal consciousness hypothesis and the access-consciousness hypothesis, Block (2007, 2011) tends to lean on arguments that take the form of an inference to the best explanation. He argues that in accounting for all the available data, we should also place substantial weight on subjects' reports about their own phenomenology. I take Block to be arguing that we should take people's phenomenology into account even in cases where the purported phenomenology reported does not give rise to downstream discriminatory or representational capacities.

Consider Block's argument relying on the famous partial-report paradigm by Sperling (1960). In this paradigm participants are presented with 3 rows of 4 letters for somewhere between 50-500 ms. After the offset of these letters, participants are presented with one of three different sounds which is a cue for the participant to report the letters of one specific row, after which they may report any other letters they can recall (Sperling, 1960)[30]. Participants can often recall 3 or 4 letters on the randomly probed row but rarely any more letters than that. Block (2007) argues that the reason why a subject can report almost all the letters of a randomly probed row but no more, is proof that people were phenomenally conscious of all 12 letters across the three rows (due to the large capacity of iconic memory) but only had cognitive access to the letters of the cued row (due to the limited capacity of working memory). As such, Block (2007, 2011) argues that an explanation which can account for both subjects' capacity-limited performance (the report of 3-4 letters of any cued row) and for the plausible assumption that subjects experienced all 12 letters of the stimulus, is taken to be better than an explanation that accounts for people's discriminatory capacities but which does not account for how subjects could have a conscious experience of all 12 letters. Block (2007, 2011) takes this to show that the better explanation is the one that assumes that phenomenally conscious states can overflow our access-conscious states[31]. That is, one must appeal to merely phenomenal conscious perceptions to explain the discrepancy between the apparent richness of people's phenomenology of the 12 letters and the relative sparseness of their discriminatory capacities, thus vindicating the merely phenomenal consciousness hypothesis.

Importantly, Block's position here does not require endorsing the extreme, empirically intractable version of this hypothesis, described above. His arguments presuppose that phenomenal states can at least partially inform introspective judgements and thus can inform the relevant phenomenological criterion of an inference to the best explanation. The challenge here is not whether some causally inert phenomenal states exist, but whether our best empirical and methodological tools reveal a systematic mismatch between the phenomenology subjects report and the representational capacities they demonstrably possess.

---

[30] In a full-report paradigm, participants are simply asked to report as many of the 12 letters as possible. In this case participants can on average report 4.3 letters.

[31] By, contrast, a series of authors have interpreted this data differently (Cohen & Dennett, 2011; Kouider et al., 2010; Phillips, 2011; Stazicker, 2011). They argue that while people do experience more than the letters they can report, people only see generic features or fragments of the unreported letters, and not rich determinate letters.

An empirical study by Cova et al. (2021) provides a great example of how we can make progress in this way. Their study directly challenges Block's conclusion on his own terms. In this paper, they both discuss the original studies that Block's argument is based, and also replicate the Sperling experiments. First, they found that the original studies did not have a methodological protocol for eliciting reports about people's phenomenology, which meant that Block's appeal to participants' rich phenomenology may be unfounded or at best based on an assumption (there were no reports recorded about whether people seemed to experience all 12 letters). Second, when they replicated the studies using an extensive protocol for eliciting reports about people's phenomenology, they found that participants took their experience of the 12 letters as being generally vague, generic, or fragmented and on all accounts described their experience to be less rich and detailed than required for Block's argument to go through[32]. As a result, the study seemingly undermines Block's conclusion that the best explanation of people's phenomenology and discriminatory capacities is that phenomenal consciousness overflows access-consciousness. In fact, the results indicate that the phenomenal overflow explanation describes people's phenomenology as richer than it is according to their own descriptions. Block as such faces the issue of explaining why people report their phenomenology as being vague or generic.

I think that a lesson to draw from this is that part of the conflict between competing views on how phenomenology and attention relate and between the different views on consciousness, may arise due to limitations in how psychological studies collect phenomenological data. This is a critique that has also been raised elsewhere in psychology itself (see Haun et al., 2017). Much psychophysical research favours highly controlled button-press paradigms (or equivalently restricted report paradigms) that serve to minimise noise but also restrict the kinds of

---

[32] Cova et al. (2021) conducted the experiment in three blocks, each consisting of: (i) a Sperling task, (ii) a questionnaire about participants' experience, and (iii) a follow-up interview. In the Sperling task, participants reported an average of 4.2 letters correctly (consistent with the original findings). The detailed responses show that only 4–8% reported seeing all 12 letters in detail and being able to identify them, and 8–14% reported seeing all 12 letters in detail without being able to identify them. Most participants reported seeing all 12 letters but not in detail (34–38%) or seeing most letters (26–36%), while 12–14% reported seeing only small parts. Overall, participants tended to describe their experience as lacking detailed, identifiable content; many reported perceiving vague or fragmented letters rather than fully detailed ones.
In interviews, participants freely described their experience and then indicated which of two hypotheses best matched it. These included the phenomenal overflow explanation (by Block) or the generic/fragmented letters hypothesis (that people at best experienced generic features or fragments of the unreported letters, see Cohen & Dennett, 2011; Kouider et al., 2010; Phillips, 2011; Stazicker, 2011). About 62% of descriptions were classified as fitting with the generic or fragmented hypothesis, about 16% as fitting with the phenomenal overflow hypothesis, and about 20% were classified as vague. When directly asked, 72% endorsed the generic phenomenology explanation, whereas only 22% endorsed phenomenal overflow. Importantly, letter-report accuracy correlated with reports of reduced phenomenal richness.

phenomenological information that can be collected (Haun et al., 2017). While they are pragmatically attractive, such paradigms may encourage a form of tunnel vision that leads us to vastly underestimate people's phenomenology rather than appropriately interrogating it, thus perhaps unjustifiably vindicating the access-conscious hypothesis. Moreover, it allows advocates of the merely phenomenal consciousness hypothesis to dig further theoretical trenches, arguing that people's experience must be richer than these arguably 'biased' studies suggest.

To avoid ignoring both the potential phenomenological richness of people's experience and theoretical trench-digging, studies investigating the relation between phenomenology and discriminatory capacities should employ more comprehensive protocols for collecting nuanced phenomenological data. These may include asking participants to select among a broad range of descriptions of their phenomenology related to the study's target hypotheses, followed by open-ended fine-grained phenomenological characterisations. It could also include free-description paradigms, in which participants describe their experiences in natural speech. These responses can then be systematically categorised in relation to competing theoretical interpretations (Cova et al. (2021) present an excellent example of this kind of research).

Applying open-ended and structured reporting methods makes it possible to gather reports about people's phenomenology that are less theoretically loaded with assumptions about what counts as conscious perception. This, in turn, allows competing views to be evaluated on more equal empirical footing. In what follows, I apply these methodological considerations to the study of temporal binding. In particular, I investigate the phenomenology of temporal binding in ways that allow for more careful and systematic exploration of how the perceptual content and phenomenology of temporal binding depend on attention.

### 2.2 Temporal Binding and Conscious Perception

Although the phenomenology of temporal binding remains relatively underexamined, I set out a provisional description of it here, drawing on resources from the philosophical literature on the unity of consciousness and on the temporal structure of conscious experience. This will make the target phenomenon of this chapter clearer.

To begin with, the phenomenology of temporal binding is not ordinarily experienced as a local or piecemeal phenomenon. Rather, when we attend to our perceptual experience as it unfolds, sensory features appear (across the perceptual field) to be presented together in time, as belonging to a single unified experiential moment. Colours, motions, sounds, and tactile sensations are not merely experienced simultaneously. They are experienced as temporally integrated with one another, as unfolding together within what seems like a shared temporal frame. This description aligns with the notions of temporal binding unity, comparability, and timeline unity introduced in Chapter 4.

We can further describe this phenomenology by borrowing the descriptions offered by various philosophers under headings such as co-consciousness, synchronic unity, and phenomenal unity (Bayne, 2010; Bayne & Chalmers, 2003; Dainton & Bayne, 2005; Prinz, 2012; Rashbrook, 2013b). Across these views, conscious experience is not characterised as a mere aggregate of independent experiential parts, but as a unified phenomenal field. As Bayne (2010) emphasises, phenomenal unity is not simply the fact that experiences occur at the same time, but that they are experienced together, as parts of a single conjoined conscious episode.

Dainton's (2000) notion of synchronic unity also captures parts of the phenomenology at issue. On his account, synchronic unity consists in a relation of co-presence according to which individual experiential contents are directly experienced together as part of a single unified phenomenal moment.

We can use these descriptions to cash out the apparent phenomenology of global temporal binding as the phenomenology of experiencing all perceptual features of our perceptual field as temporally anchored within a single unified conscious experiential moment, such that all these features are experienced as temporally related to one another.

However, the apparent phenomenology of global temporal binding raises questions central to this chapter. What is the nature of this apparent phenomenology of global temporal unity? Is it the case that perceptual features across the entire perceptual field are genuinely experienced as temporally bound? Or is the phenomenology of temporal binding more local, partial, or indeterminate than this description of our apparent phenomenology suggests?

To sharpen this question, it is crucial to distinguish between: (a) the phenomenology of temporal binding, understood as what it is like, subjectively, to experience sensory features as temporally bound, and (b) the content of temporal binding, understood as the temporal relations between features that are represented or encoded in perceptual states in a way that supports discrimination, report, or action.

This distinction parallels the broader debate, discussed above, between advocates of the access-consciousness hypothesis and the phenomenal-consciousness hypothesis. One might endorse the access-consciousness hypothesis and claim that the phenomenology of temporal unity is fully determined by representational content, such that to experience features as temporally bound just is to perceptually represent their temporal relations. Alternatively, one might endorse the phenomenal-consciousness hypothesis and allow for a gap between phenomenology and representational content, such that the phenomenology of temporal binding may overflow what is represented or accessed at any time.

This distinction bears directly on the role of attention in temporal binding. If the phenomenology of temporal binding is reducible to, or exhausted by, representational content, and if such content depends on attention, then temporal binding phenomenology will likewise be attention dependent. By contrast, if the phenomenology of temporal binding is not reducible to attention-dependent representational content (if, for example, it is grounded in pre-attentive processes that give rise to the experience of a globally temporally unified perceptual field), then it is plausible that temporal binding phenomenology overflows attentional capacities.

The former position fits naturally with my proposed hybrid theory of temporal binding, presented in Chapter 3. According to this view, attention is necessary for engaging in temporal binding, and both the content and phenomenology of temporal binding are constrained by attention-dependent perceptual representational capacities. In this respect, the hybrid theory aligns with the access-consciousness hypothesis. Conscious perceptual experiences of temporal binding depend, at a minimum, on attention, and there are no merely phenomenal experiences of temporal binding that outstrip what can be perceptually represented and accessed.

The latter position fits naturally with the brain time theory discussed in Chapter 2. According to this theory, temporally bound experiences are determined by the time at which sensory information ascends the sensory processing hierarchy and reaches perceptual endpoints, understood as the ‘requisite processes responsible for conscious experience’ (Arstila, 2016a). If experiences of temporal binding arise directly from simultaneity at these perceptual endpoints, and if reaching such endpoints is not attention-dependent (as is assumed by Arstila (2016a) and Zeki[33] (2007; 2015)), then attention is not necessary for temporal binding. Combined with the

---

[33] Note while Zeki (2007, 2015) supports that the experience of sensory features can be pre-attentive, he assumes that sensory features are not experienced as a unified whole (Zeki, 2003), in the sense described by for example Bayne’s Unity Thesis (2010). However, Zeki still takes synchrony in sensory processing to be sufficient for an experience of temporal binding, in the sense that co-processed features are experienced as occurring simultaneously (but without any experiential unity connecting these individual co-occurring experiences). Thus Zeki’s view is compatible with a unity-free global phenomenology of temporal binding according to which we simply experience all co-processed sensory features as occurring simultaneously. However, it is likely not a view he would sign up to, due to his beliefs about the asynchrony of perceptual processing (Zeki, 2015).

phenomenal-consciousness hypothesis, this opens the possibility that we have merely phenomenal experiences of temporal binding that are pre-attentive and global.

While these two views may agree that perceptual representations of temporal binding (those that support discrimination, report, and action) depend on attention, they disagree about what follows for phenomenology. This disagreement gives rise to two competing positions:

> ***Attention-dependent phenomenology of temporal binding:*** Attention is necessary for conscious perceptual phenomenology of temporal binding.
>
> ***Pre-attentive phenomenology of temporal binding:*** Attention is not necessary for conscious perceptual phenomenology of temporal binding.

The attention-dependent view motivates a deflationary interpretation of the apparent phenomenology of global temporal binding. If attention constrains both temporal binding content and phenomenology, then any discrepancy between the apparent global character of temporal binding and the limits of attentional capacity need not reflect a genuine mismatch between phenomenology and the representational contents of our perceptions. Instead, it may result from systematic subjective inflations that lead us to overdescribe our phenomenology of temporal binding as richer and more global than it in fact is.

By contrast, the pre-attentive view motivates a non-deflationary interpretation of this phenomenology. If temporal binding is grounded in pre-attentive processes, and conscious phenomenology can be pre-attentive, then conscious experiences of temporal binding need not be constrained by attentional limits, and it becomes plausible that perceptual features across the entire perceptual field are genuinely experienced as temporally bound (despite the fact that our discriminatory capacities are constrained).

The focus of the rest of this chapter is to develop a framework for how we can empirically study which of these interpretations of the phenomenology of temporal binding is better supported. Empirical findings based on this framework would put the claims of the hybrid theory and brain time theory of temporal binding to an empirical test.

## 3 On Empirically Investigating the Relation Between Attention and Temporal Binding

In the following sections, I critically examine the attention-dependent phenomenology of temporal binding view and the pre-attentive phenomenology of temporal binding view by investigating three central questions and related empirically testable hypotheses. Question 1 concerns how the richness and sparseness of our phenomenology of temporal binding depend on attention (§3.1).

Question 2 concerns whether there is a discrepancy between how the phenomenology of temporal binding and our temporal binding discriminatory capacities depend on attention (§3.2). Question 3 concerns how to interpret any such (potential) discrepancy (§3.3).

### 3.1 Question 1

To investigate the relation between our phenomenology of temporal binding and attention, we must answer the following question (Q1): Does the apparent richness of people's phenomenology of temporal binding depend on attention?

But, to answer Q1 we need to first define the terms 'richness' and 'sparseness' relative to the phenomenology of temporal binding. The distinction between rich and sparse views of perceptual experience provides the background for understanding the competing positions on temporal binding phenomenology.

#### *3.1.1 Is Perceptual Experience Rich or Sparse?*

Closely related to the dispute between the single vs. double dissociation views and the access- vs. merely phenomenal consciousness hypothesis is a long-standing debate about whether perceptual experience is rich or sparse.

Some theorists defend a sparse view of perceptual experience, arguing that perceptual phenomenology depends on attention, working memory, and other capacity-limited cognitive resources (Baars et al., 2013; Baars, 1988; Cohen & Dennett, 2011; Dehaene, 2014; Dennett, 1991; Lau & Rosenthal, 2011; Marchetti, 2012; O'Regan & Noë, 2001). On this view, conscious perceptual experience is constrained by what perceptually processed features can be attended to and cognitively accessed.

Others defend a rich view of perceptual experience, claiming that pre-attentive early cortical processing can itself give rise to conscious experiences (Aru & Bachmann, 2013; Block, 2011; Bronfman et al., 2014; Iwasaki, 1993; James, 1890; Koch & Tsuchiya, 2007; Lamme, 2003, 2006; Siewert, 1998; Zeki, 2003). According to this view, phenomenology can overflow what perceptually processed features can be attended to or cognitively accessed, yielding conscious experiences that extend beyond attentional limits.

As such, this dispute maps neatly onto the distinction between single dissociation view and the access-consciousness hypothesis (typically associated with the sparse view) and the double dissociation view and merely phenomenal-consciousness hypothesis (typically associated with the rich view). The paradigmatic disputes are about colour vision and natural scene perception.

The rich view argues that we have rich conscious colour experiences across our entire visual field, and that our perceptual experiences of natural scenes extend beyond the parts we attend to (Block, 2014; Chalmers, 2006; Haun et al., 2017; Koch & Tsuchiya, 2007; Li et al., 2002; Rensink, 2000; van Boxtel et al., 2010; Zeleznikow-Johnston et al., 2023) On this interpretation, our description of our apparent rich phenomenology is taken at face value, as reflecting something true about what our conscious experiences are like. The rich view seems to satisfy many intuitions about our experiences. It appears as if we perceive the world in much more detail than suggested by views who argue that our perceptual experience depends on attention. We seem to experience things we do not directly attend to. The rich view for example speaks to our intuition that we have rich conscious colour experiences across our entire visual field despite being unable to accurately report all colours in our visual periphery (Block, 2014; Chalmers, 2006; Haun et al., 2017; Zeleznikow-Johnston et al., 2023). Moreover, we seem to have rich experiences of natural scenes. The perceptual experience of other objects as well as the background features of a scene does not change, disappear, or become blurry/vague when and if we attend to certain parts or objects of that scene.

By contrast, the sparse view argues that our peripheral colour experiences and scene perceptions are sparser than the way we tend to describe them (Abramov et al., 1991; Cohen et al., 2016; Cohen & Rubenstein, 2020; Dennett, 1991; Gordon & Abramov, 1977; Hansen et al., 2009). The sparse view, for example, relies on experimental evidence of our poor capacity to discriminate colours in our peripheral vision to argue that our peripheral colour experiences are sparse or non-existent despite how they may appear (Abramov et al., 1991; Cohen et al., 2016; Cohen & Rubenstein, 2020; Dennett, 1991; Gordon & Abramov, 1977; Hansen et al., 2009). They argue that this impoverished experience is primarily caused by two things. First, colour perception varies across the visual field, being best in the fovea and getting gradually worse in the intermediate periphery. Red-green colour sensitivity for example steeply declines in the intermediate periphery compared to blue-yellow colours and is almost absent in peripheral vision above 40 degrees from the central fixation point (Hansen et al., 2009). Second, colour experience and perceptual experience are dependent on attention, such that our minimally attended peripheral vision is greatly impoverished (as long as our eyes do not saccade putting peripheral sensory features into more foveal and attended vision) (Cohen et al., 2016; Cohen & Rubenstein, 2020; Gordon & Abramov, 1977; Hawkins et al., 2022).

According to the sparse view, the connection between attention and perceptual experience also holds for our perception of natural scenes (Cohen et al., 2011, 2021; Li et al., 2002). This leads to the strong claim that our experience of natural scenes becomes greatly impoverished or even absent when some foveal perceptual tasks fully consume our attention. As Cohen et al. (2021) say:

> 'While it is easy to imagine a person failing to notice the keys on the dining room table, it is difficult to imagine how they could fail to notice that their dining room has been changed into a glacier'

Despite this being a very strong claim, Cohen et al. (2021) take themselves to have experimentally shown that when our attention is fully consumed we do not even notice large-scale changes in our peripheral visual field.

This dispute over the phenomenological richness of natural scene perception is closely connected to work on gist perceptions. Gist perceptions are a kind of perception that represent the summary statistics of a scene. For example, we can represent a scene as a forest without first paying individual attention to a number of trees that makes it reasonable to perceptually judge the scene as a forest (Greene & Oliva, 2009). We can call the information that is picked up by gist perceptions 'ensemble statistics' (see Cohen et al., 2016), because they pick up statistical relations of ensembles of various features.

Wolfe et al. (2011) suggest that gist perceptions rely on a fast visual processing pathway to represent ensemble statistics that is distinct from the slower selective visual pathway guided by semantic and episodic information, leading to a bottlenecked process, gist perceptions may rely on a non-selective pathway that is not subject to the same constraints (Wolfe et al., 2011). This non-selective pathway is good at what Wolfe et al. (2011) call 'Global non-selective image processing' which they argue 'allows observers to rapidly extract statistical information from the entire image [of a scene]. Observers can assess the mean and distribution of a variety of basic visual feature dimensions.'

There is ample evidence showing that gist perceptions allow for the conscious awareness of the distributions of basic features such as colour and texture (Rousselet et al., 2005), the spatial layout of scenes (Oliva & Torralba, 2001, 2006), the mean and distribution of different basic visual features such as size (Chong & Treisman, 2005), general orientation (Dakin & Watt, 1997), average velocity and direction of motion (Watamaniuk & Duchon, 1992), the semantic categories of a scene (Greene & Oliva, 2009; Groen et al., 2013b), and the general magnitudes of features (Demeyere et al., 2008).

The mechanisms underlying gist perceptions offer a potential explanation for why perceptual experiences appear richer than we would expect if our conscious perception relied solely on the slower and more constrained selective attentional mechanisms.

The sparse and rich views disagree about whether gist perceptions are best interpreted as supporting the idea that we genuinely enjoy rich phenomenology and whether such gist perceptions require attention. While proponents of the rich view take gist perceptions to provide rich, attention-independent experiences (Bronfman et al., 2014; Koch & Tsuchiya, 2007; Levin & Simons, 1997; Sampanes et al., 2008), proponents of the sparse view take gist perceptions to be attention-dependent (Cohen et al., 2011, 2016, 2021; Cohen & Rubenstein, 2020).

One example of a study that is taken to support the rich view comes from Bronfman et al. (2014). They claim that we can perceive colour diversity 'cost free' meaning that detecting this feature does not depend on or require attention. They claim that after participants are shown a Sperling-like display of coloured letters (3 rows of 4 letters), where one row is pre-cued, they could estimate the diversity of colours of the rows that weren't cued, without this incurring any cost on their capacity to report the letters. They take this to suggest 'that color diversity is registered automatically, outside focal attention, and without consuming additional working memory resources.' (Bronfman et al., 2014, 1394).

However, other studies challenge this. Brady and Tenenbaum (2013) for example tested participants' ability to indicate how many of 25 items change colour (from red to blue). They show that in a condition where people can only attend selectively to a few items at a time, they perform equally well regardless of how the statistical configurations of the display were structured. In another condition where people can engage in gist perceptions to take note of the overall structure of the display of colours, performance increases vary as statistical regularities are manipulated (increasing the statistical regularities of the display of colours increased the number of colour changes stored in visual memory from ~4.5 items to ~24 items). This shows, as Cohen et al. (2016) argue, that people's ability to discriminate colour changes via gist perceptions is a function of how well they can track higher-order regularities of the display.

So, if gist perceptions of colour truly 'came for free', we should expect this gist perceptual experience of colour variations to remain unaffected by intervention in how attentionally demanding it is to pay notice to the higher-order structure of some scene. However, this is not the case. Hawkins et al. (2022) similarly show how our experience of colour diversity in peripheral vision depends on attention.

In fact, many studies seem to support how gist perceptions are attention dependent: the attentional blink and dual-task paradigms affect natural scene perception even when the main target in the central visual field is perceived and the scene target is neurologically processed (Marois et al., 2004; Slagter et al., 2010), attention-demanding dual-task paradigms affect gist-like scene perception (Rousselet et al., 2005; Stein et al., 2009), and lastly, inattentional blindness (phenomenal blindness due to demanding and distracting attention tasks) seems to occur for natural scenes (Cohen et al., 2011).

Taken together, these studies suggest that even gist perceptions have an attentional cost and that statistical ensembles can go phenomenologically undetected if our attention is fully consumed (Cohen et al., 2021). If this is correct, then there may be no need to posit non-accessed, merely phenomenal conscious experiences to explain our phenomenology originating from gist perceptions. The phenomenology of gist perceptions can be explained as a function of how attention and working memory capacities relate to our ability to track the higher-order statistical regularities of the perceived ensemble statistics.

Importantly for the purposes of this chapter, the fact that gist perceptions only give us conscious perceptual information about higher-order statistical information also suggests that gist perceptions may not be particularly helpful when it comes to accounting for the phenomenology of the multimodal integration of sensory features of objects, i.e., spatial and temporal binding.

Gist perceptions are unlikely to underpin our ability to spatially and temporally bind features into coherent, distinct spatiotemporal objects and events. Spatial as well as temporal binding require the selective attention of features of objects such that they can be encoded as occurring at a specific spatiotemporal location. In the case of spatial binding, it requires that objects have a determinate retinotopic location across different feature maps. Temporal binding, as I have argued in Chapters 3 and 4, requires that attention selectively modulates specific sensory features in the world and temporally integrates them relative to their best temporal fit, given dynamical internally and externally biased top-down and bottom-up demands. The global non-selective image processing of gist perceptions, which only gives us access to global statistical ensembles of a scene (Wolfe et al., 2011), supports neither of these. That is, I doubt that we can use these mechanisms to bind sensory features into objects with determinate spatial and temporal locations.

Of course, ensemble statistics may be necessary and useful for guiding action in various ways, but they do not guide us in the same way as perceptually temporally and spatially binding features do. Nor do they provide us with the same kinds of perceptions.

For example, perceiving the general orientation, speed, and colour of cars on a road is enough for me to determine the direction of traffic, the general speed, and the general colour. However, it does not give me a perceptual experience of the specific speed, colour, or orientation of the individual cars. Similarly, when I enter a crowded room and immediately register that the overall noise level is rising, I may raise my voice in response. This adjustment is guided by a statistical summation, the average amplitude of speech in the room, yet it does not require, nor does it provide any perceptually bound information about the loudness of particular speakers or moments of utterances.

In the next sections, I suggest that while gist perceptions may give us access to some temporal information about minimally attended and peripheral features, it is not sufficient for the temporal binding of such features.

### *3.1.2 Rich vs. Sparse Temporal Binding*

The dispute between rich and sparse views of perceptual phenomenology provides the foundation for understanding the two competing views of temporal binding phenomenology set out in §2.2. Much like the sparse view of perceptual experience, the hybrid theory predicts that our phenomenology of temporally bound perception is local and thus sparse rather than global and rich.

These considerations allow us to set out more specific versions of the 'Attention-dependent phenomenology of temporal binding' and 'Pre-attentive phenomenology of temporal binding' set out above:

> ***Attention-dependent phenomenology of temporal binding:*** We can only have phenomenology as of determinately temporally bound objects when we selectively attend to sensory features. Gist perceptions support, at best, indeterminate, ensemble-level phenomenology of temporal relations. We have no phenomenology of temporal binding for unattended sensory features.
>
> ***Pre-attentive phenomenology of temporal binding:*** We can have phenomenology as of determinately temporally bound objects regardless of whether these perceptions rely on selective attention, gist perception, or unattended processing.

These positions are tested via hypothesis 1:

> **H1:** The richness of people's phenomenology of temporal binding is affected by the degree to which they can attend to the relevant features. The less they can attend to the relevant

features to be temporally bound, the less rich and determinate their experiences of temporal binding are.

H1 is intended to capture the core commitment of the attention-dependent temporal phenomenology of binding view. If H1 is false, so is the 'Attention-dependent phenomenology of temporal binding' view. In this case we should expect people's phenomenology to remain relatively stable regardless of how attention is manipulated. If H1 is true, then 'Attention-dependent phenomenology of temporal binding' is likely on the right track and we should observe a necessary and systematic variation in phenomenological richness as a function of attentional load. Temporal binding phenomenology should be subject to standard attentional effects such as inattentional blindness, the attentional blink, and effects of dual-task interference which are standard paradigms used for studying the relationship between conscious experience and attention (Cohen & Chun, 2017). Critically, H1 concerns phenomenology directly, not performance. Even if accuracy tracks attentional demand, we must independently establish what participants experienced as temporally bound.

Importantly, H1 does not predict that minimally attended regions of the perceptual field are phenomenally empty. It allows that observers may have temporal experiences in our minimally attended peripheral vision that represent gist perceptual ensemble statistics. What it denies is that these ensemble statistics represent features as being temporally bound. That is observers would not experience colour–motion animations processed via gist perceptual pathways as determinately temporally bound.

Testing H1 therefore requires experimental paradigms that directly probe people's apparent phenomenology of temporal binding under varying attentional demands, rather than deriving this solely based on task performance. To this end, I propose a series of tasks intended to be templates for future empirical studies (§3.1.2 below) and in §3.1.3 I present a protocol for how to elicit reports about people's phenomenology in these tasks.

#### *3.1.2 A Series of Template Animations To Test H1*

Consider first the task illustrated in Animation 2a[34]. Participants would here be instructed to fixate on the central cross and to note how many of the individual animations they simultaneously experience as having determinate colour–motion pairings (having an experience of which colour is co-occurrent with which motion direction).

---

[34] See all animations here: https://rpthesis.wordpress.com/

Upon introspection, I find it difficult to attend to more than three or four, at most five, animations at once without losing track of the experience of the temporal binding of some of the colour–motion pairs. Even when it seems that several pairings are experienced as temporally bound, it is unclear whether this reflects a single simultaneous perceptual experience or a rapid shifting of attention combined with short-term memory and inferences about what the experience of the previously attended animation was like. Moreover, when attention is focused on a limited subset of the animations (e.g., ~3 pairings), the pairing of remaining animations are either experienced as indeterminate or not experienced as bound at all. Instead, they appear merely as flashes of colour accompanied by some unspecified motion.

This pattern of experience is precisely what H1 predicts. If attention allows access to only a limited number of colour and motion transients at once, then H1 entails that only those features can be experienced as temporally bound. The unattended pairings should not be experienced as bound, even if they are simultaneous in event time. For H1 to be false, it would need to be the case that participants experience more colour–motion pairings as temporally bound than they can attend to. Crucially, what matters here is not performance, but participants' reports of what they experienced as temporally bound.

Importantly, H1 does not imply that minimally attended regions of the perceptual field are phenomenally empty. Rather, it allows that observers may experience coarse or generic temporal structure in peripheral vision. What H1 denies is that observers experience specific colour–motion pairings in these regions as determinately temporally bound. In line with the sparse temporal binding view, minimally attended perception may support gist-based or ensemble representations that capture general patterns of change without encoding specific feature pairings. The hypothesis simply says that we do not experience the minimally attended animations as temporally bound in the rich way we experience the attended animations as temporally bound.

This contrast becomes vivid when comparing your experience of Animation 5a to your experience of Animation 2b. In Animation 2b, all the animations conform to a uniform statistical structure. As a result, the animations can be experienced at a gist as a single fragmented gestalt exhibiting a shared temporal pattern. These are experiences as such despite the fact that our selective attention cannot be devoted to each element individually. If one animation were inverted, it would be easy to detect, precisely because it would violate the ensemble statistics of the display provided by our gist perceptions. In this case, temporal binding across the field is experienced as unified not because attention tracks each pairing, but because the perceptual system can rely on a single, coherent statistical pattern represented at a gist.

This comparison supports H1. Although Animations 2a and 2b contain the same amount and number of changes, the associated phenomenology differs markedly. This seems to be both due to the fact that pairing the features of the Animation 5a demand more attention than in the case of Animation 5b and because the animations in Animation 5b conform to a uniform statistical pattern that can be represented at a gist. In this way Animation 5b allows an experience of temporal binding of all the animations because this experience is based on ensemble statistics. It thus seems that the difference in phenomenology tracks differences in attentional demands, lending support to the sparse temporal binding view.

A further test of H1 can be constructed by placing a large, attention-demanding colour–motion animation at fixation, surrounded by smaller peripheral animations, some of which are inverted (see Animation 3). If H1 is correct, attending to the central animation should substantially reduce the richness of temporal binding phenomenology of the animations in the periphery. This would occur because the central animation consumes attentional resources and interferes with the perceptual interpretation of peripheral temporal relations.

Change blindness paradigms offer another way to test H1. Change blindness paradigms involve presenting a participant with two visual scenes with a brief intervening screen in between. In successful change blindness paradigms, people fail to notice or experience changes between two visual scenes due to a failure to attend to the change. Change blindness paradigms are useful because they form a reliable paradigm for testing the role that visual attention plays in perceptual experiences (Such paradigms are often appealed to in this literature, see Block, 2007; Brady et al., 2009; Busch, 2013; Cohen et al., 2012; Cohen & Chun, 2017; Dennett, 1992; Dretske, 2007; Landman et al., 2003; Mack & Rock, 1998; Noë et al., 2000; Rensink et al., 1997; Simons, 2000).

Animation 4 is a classical change-blindness paradigm adapted to the case of temporal binding. For this animation participants are to fixate on a cross and observe a display of colour–motion animations for approximately two seconds, followed by a brief blank screen and then a second presentation (the whole thing is then repeated). In the last scene, individual orientations of the animations may or may not have changed (in this specific case, the orientation of one is different). The task is for the participant to state whether they noticed any change and what that change is.

By relying on a Sperling paradigm (Cova et al., 2021; Reeves & Sperling, 1986; Sperling, 1960) with a change-blindness element developed by Landman et al. (2003)[35], we can modify Animation 4 to construct a temporal binding change blindness task that has a few different conditions intended to tease out the extent to which observers rely on attention-dependent or pre-attentive visual mechanisms to encode information about the perceptual targets in the first scene.

Animation 5 illustrates three versions of Animation 4 that differ only in when participants are cued to the potentially changing animation. They are either cued during the initial stimulus (5a), during the blank screen (5b), or during the final stimulus (5c). In this specific setup, the task is to tell whether a specific cued animation was inverted between the first and the last scene.

Conditions (5a) and (5b) are especially important. This is because performing consistently well in (5a) requires maintaining the temporal binding of only one animation, whereas performing consistently well in (5b) requires maintaining the temporal binding of all animations until the cue appears. The only way to perform equally well in (5b) as in (5a) would be to retain information about the temporal binding of all animations in a way that does not depend on attention or working memory.

If the performance for condition (5a) and condition (5b)is consistently the same across several trials, then this might mean that participants continue to maintain some phenomenologically visual representation of the whole array of temporal binding of the colour-motion animations. This is important because perceptually experiencing all the feature pairs of all the animations as temporally bound and retaining this information to make a report about the cued animation (regardless of whether they at this point can report anything about the pairing of the other animations) would mean that their phenomenology of temporal binding outstrips the capacities of working memory and attention. So if the performance profile is the same for (a) and (b), then it may be that the pairings of all the colour-motion animations is phenomenally experienced despite all the pairings not being access-conscious. All the pairings do obviously not constitute access-conscious perceptions, because participants can only tell whether the (randomly) probed colour-motion pairing has changed, not whether the non-probed colour-motion pairings have changed.

---

[35] In Landman et al. (2003) subjects fixate see differently oriented rectangles (500 ms), followed by a grey screen (200-1500 ms), followed by the initial stimulus of rectangles which individual orientations may or may not have changed. Participants are shown one of three conditions, where conditions differ in whether the cue is given in the first, second, or third scene. The task is for participants to state whether the cued rectangle has changed orientation or not.

This line of reasoning mimics Block's (2007) use of Landman et al. (2003) to argue for the hypothesis that the merely phenomenal consciousness hypothesis best explains the rich phenomenology of the participants (mentioned in §2). Translating the argument to temporal binding, one might argue as follows:

> ***Phenomenological overflow hypothesis:*** In scene 1, observers have a clear conscious perceptual experience of the temporal binding of all the initial colour motion pairings, despite not being able to identify all of them. Their phenomenology overflows what they can consciously represent which is why participants can, at most, report the feature pairing of the cued animation.

I think that the phenomenological overflow hypothesis can be proven wrong, much like Block's (2007, 2011) phenomenological overflow hypothesis in the case of visual shape perception has been proven wrong by Cova et al. (2021) (see §2). Following Cova et al. (2021), Cohen & Dennett (2011), and Stazicker (2011) arguments against Block, I suggest that the following is right:

> ***Generic phenomenology hypothesis:*** In scene 1, observers consciously experience temporal relations for more features than they can discriminate or report, but this experience is as of generic and indeterminate, temporal relations consisting of ensemble statistics of motion and colour changes, rather than temporally bound feature pairings. The cue is what cause the experience of temporal binding of the feature pairs. The cue cause us to attentionally modulate one set of unconsciously represented features so that these become temporally bound before the unconscious representation of these features decay.

An indication that something like the generic phenomenology hypothesis may be right can be brought out by comparing the change-blindness case in Animations 5a, 5b, and 5c to the ones in Animations 6a, 6b, and 6c. In Animations 6a, 6b, and 6c all the colour-motion patterns are identical in the initial stimulus. That is, the colours and motion directions of the animations are the same in scene 1. So in scene 3, where the cued animation might have been inverted, the inversion of the animation would completely violate the general pattern of the ensemble statistics of the animations presented in scene 1.

As a result, it is much easier, in all three conditions, to get the right answer about whether the cued animation is inverted or not. We should not expect there to be any discrepancies in performance on tasks like Animations 6a and 6b, even if there is a large discrepancy between performance in tasks like Animations 5a and 5b. This is because in Animation 6 the ensemble structure is uniform, and so we can rely on selective attention and gist perceptions to perceive all the colour motion patterns as a single gestalt. This makes it very easy to perceive if one of the pairings has been inverted or not. As such, observers' performances for all conditions in Animation 9 would probably all be more or less identical. However, once there is no clear pattern to latch on to, as in Animation 5, it becomes much more difficult to perceive the temporal binding of the different colour-motion pairs. That is probably because our attention is completely overwhelmed in the case of Animation 5 and far from overwhelmed in the case of Animation 6. The generic hypothesis would predict something like this difference in performance between Animation 5 tasks and Animation 6 tasks because the difference in phenomenology and performance both follow differences in attentional demand.

Finally, as I mentioned above, it is important to emphasise that performance alone cannot settle these issues. Even if accuracy tracks attentional demand, we must also elicit detailed phenomenological reports concerning how rich or sparse participants' experiences of temporal binding seem to them. Such reports are crucial both for assessing H1 because such reports may both co-vary with expected performance, and they may not. This also means assessing the difference between people's apparent phenomenology in Animation 5 and Animation 6. If people report the same phenomenology across these figures, then that is at least preliminary evidence that there is a large discrepancy between representational capacity and phenomenology. This data is essential for assessing 'Question 2' (§3.2), whether or not attention constrains discriminatory representational capacities and phenomenology in the same way, and 'Question 3' (§3.3), whether we subjectively inflate our phenomenology or not. Below I discuss what such phenomenological reports involve.

### *3.1.3 Asking People About Their Phenomenology*

To empirically assess whether H1 is true or false, we need methods for carefully and systematically inquiring into what people's experiences are like when they engage in different temporal binding tasks. Crucially, to avoid begging the question against the phenomenological overflow hypothesis, we cannot rely solely on task performance as a proxy for phenomenology. Button presses, accuracy scores, and reaction times tell us something about discriminatory representational capacities, but they do not, by themselves, reveal what participants experienced as temporally bound and how such experiences appeared to them. This limitation is a well-known shortcoming of many psychophysical paradigms (see Haun et al., 2017).

Accordingly, testing H1 requires collecting phenomenological data that go beyond simple performance measures, while still being sufficiently structured to allow meaningful comparison across participants and conditions. There are at least two complementary ways of doing this.

The first is to use an extended response paradigm, similar to that of Cova et al. (2021) and suggested in Haun et al. (2017). In such a paradigm, participants are asked to characterise their experiences by selecting from a broad range of descriptions that differ systematically in how rich, determinate, and attention-dependent the reported phenomenology is. For example, after completing a temporal binding task, participants could first be asked a coarse-grained question such as: ‘In general, did you feel like you could pair the colour and motion direction of all the animations on the screen?’, with response options such as ‘yes,’ ‘no,’ or ‘I don’t know.’ This initial question serves to establish a rough baseline of how globally rich participants take their temporal binding phenomenology to be.

This can then be followed by a set of more fine-grained statements designed to probe the character of that phenomenology in greater detail. Importantly, some of these statements should be naturally interpreted as supporting the negation of H1, while others should support H1. Again following the strategy used by Cova et al. (2021), statements that would count against H1 might include:

(a) ‘I felt like I simultaneously saw the colour and motion changes of all the visual animations in detail at the same time and felt that I was able to identify all or most of the pairings.’

(b) ‘I felt like I saw the colour and motion changes of all the visual animations in detail, even though I could not identify all or most of the pairings.’

Statements that support H1 include:

(c) ‘I felt like I simultaneously saw the colour and motion changes of all the visual animations, but not in detail. I mainly saw that colours and directions were changing, without being able to identify specific pairings.’

(d) ‘I felt like I simultaneously saw the colour and motion changes of most, but not all, of the visual animations, and I could not identify most of the pairings.’

(e) ‘I felt like I only saw the colour and motion changes of a small part of the visual animations simultaneously, and I could not identify most of the pairings.’

Finally, a control option should test whether the participant understood the tasks:

(f) ‘I felt like I saw nothing.’

Allowing participants to choose among these descriptions makes it possible to extract far more nuanced information about their phenomenology of temporal binding than would be possible if phenomenology were inferred indirectly from behavioural performance alone. In particular, it enables us to distinguish reports of genuinely rich temporal binding from reports that reflect only generic or indeterminate experience.

A second, complementary approach is to use a free-description paradigm, in which participants are invited to describe their experiences of attended and minimally attended temporal binding tasks in their own words. These reports can then be systematically categorised according to how well they fit competing hypotheses about temporal binding phenomenology.

In this context, two categories are especially relevant. The phenomenological overflow hypothesis claims that participants have clear and determinate conscious experiences of the temporal binding of all colour–motion pairings, even though they cannot identify or report all of them. By contrast, the generic phenomenology hypothesis holds that while participants experience more than they can accurately discriminate, their experience of unattended pairings is generic, indeterminate, or statistical in nature. For example, consisting of vague colour change, overall motion, or ensemble-level structure rather than specific feature pairings.

Free descriptions that strongly support the phenomenological overflow hypothesis would count against H1 and support the rich temporal binding view, whereas descriptions that emphasise generic, fragmentary, or indeterminate temporal relations would support H1 and the sparse temporal binding view.

Both report paradigms should also include measures of confidence, for example by asking participants how confident they are in their characterisation of their experience. This is particularly important in the present context, since attending explicitly to whether sensory features appear temporally bound is an unusual and introspectively demanding task. Confidence ratings allow us to distinguish firm phenomenological commitments from hesitant or uncertain reports, a distinction that becomes especially relevant in interpreting potential discrepancies between phenomenology and performance (as discussed further in §3.3.1).

Collecting phenomenological data in this way is essential for adjudicating between the sparse and rich temporal binding views. Only by directly examining how people describe their experiences under varying attentional conditions can we determine whether apparent global temporal binding reflects genuinely rich phenomenology or instead results from generic experience combined with systematic subjective inflation.

### 3.2 Question 2

To go beyond simply testing the rich vs. sparse temporal binding distinction introduced in §3.1 and to directly assess whether the *attention-dependent phenomenology of temporal binding view* or the *pre-attentive phenomenology of temporal binding view* (§2.2) is correct, we must investigate whether our capacity to successfully engage in temporal binding tasks and our phenomenology of temporal binding are constrained by attention in the same way.

Put differently, the question here is not merely about how rich our experience of temporal binding is, but whether the apparent richness of that experience tracks our actual perceptual capacities. This question closely parallels the central issue raised by classic change-blindness paradigms, in which people's reports suggest a potential discrepancy between the amount of visual detail they can accurately discriminate and the amount of visual detail they take themselves to consciously experience. In such paradigms, viewers often describe their visual experience as phenomenologically rich and detailed across the entire scene, while nevertheless failing to notice large, salient changes to important elements of that scene (Simons & Ambinder, 2005).

The core issue here is whether an analogous discrepancy arises in the case of temporal binding. Specifically, we want to know whether people's reports about the temporal unity of their experience outstrip what their performance on temporal binding tasks indicates they can perceptually discriminate.

We can formulate this issue as follows:

> ***Q2:*** Is there a discrepancy between how (a) the apparent richness of our phenomenology of temporal binding and (b) capabilities to perform successfully on temporal binding tasks, depend on attention?

To answer this question, we can attempt to confirm or disconfirm the following hypothesis:

> ***H2 (No phenomenal overflow):*** There is no discrepancy between how people's descriptions of their experiences of temporal binding and their performance on temporal binding tasks depend on attention.

H2 should be understood as a denial of phenomenal overflow in the case of temporal binding. According to this hypothesis, people's phenomenological reports closely track their actual perceptual capacities, such that what they say about their experience of temporal binding mirrors what they can in fact discriminate, retain, or use to guide action. If participants report that particular sensory features are experienced as temporally bound in a determinate way (for example, that a red pattern is paired with leftward motion and a green pattern with rightward motion), then (if H2 is correct) this report should be borne out by their performance on the corresponding temporal binding task.

If H2 is false, then people will systematically describe their experiences of temporal binding as richer or more determinate than their performance warrants. In that case, participants may be confident that they experienced certain sensory features as temporally bound into a coherent perception, even though their performance indicates that they were unable to perceptually bind those features. This would constitute a discrepancy that calls for explanation, a task taken up in §3.3.

As such, H2 is a strong hypothesis. Any reliable intervention that reduces temporal binding performance via attentional manipulation, without producing a corresponding degradation in phenomenological reports, would count as evidence against it.

If H2 is true, then the results bearing on Q2 should show that people's discriminatory representational capacities and their phenomenological reports of temporal binding co-vary in a systematic way. Decreases in performance due to attentional distraction, overload, or divided attention should be accompanied by corresponding decreases in the reported richness or determinacy of temporal binding experience. Likewise, improvements in performance should be accompanied by richer phenomenological reports, mutatis mutandis.

If H2 is false, by contrast, then the results of Q2 should reveal a dissociation between performance and phenomenology. People would continue to report phenomenologically rich experiences of temporal binding across their perceptual field even in conditions where their performance indicates that attention was insufficient to support successful temporal binding. Such a pattern would support a form of phenomenal overflow and thereby favour the pre-attentive phenomenology of temporal binding view (§2.2) over the attention-dependent alternative.

### 3.3 Question 3

If H2 is false, then we must understand why these discrepancies occur. It may be that our phenomenology of temporal binding is pre-attentive and genuinely overflows our capacities to perceptual representations that we rely on to make temporal discriminations. However, it may also be that any such discrepancies are due to other factors, such as people subjectively inflating the richness of their descriptions due to perceptual expectancies, inferential biases, or other cognitive factors.

Subjective inflation refers to the phenomenon of overstating the richness of our phenomenology. Subjective inflation can occur in various perceptual situations and for various reasons (see Knotts et al., 2019). In this section, I present four different ways one may subjectively inflate one's experience of temporal binding. I take such subjective inflations to be due to (i) a lack of noticing when our experiences are not temporally bound, (ii) inferences based on expectations of how things would look if we were to shift our attention, (iii) inferences biased by our currently attended or central representations, and (iv) inferences based on expectations given prior knowledge of how features tend to be temporally related. On the basis of this, I propose the following hypothesis.

> ***H3 (Subjective inflation):*** People tend to subjectively inflate their phenomenology of experiences of temporal binding due to factors (i)–(iv).

If temporal binding does not occur independently of cognitive access, because temporal binding requires the active involvement of attention-dependent mechanisms, then we would expect something like H3 to be true (conditional on H2 being false)[36]. If H3 is true, we should adopt a cognitive error theory about our apparent phenomenology as of global temporal binding.

> ***Cognitive error theory:*** Our phenomenology of temporal binding is not in fact an experience as of global temporal binding. Rather, we are subject to systematic cognitive errors that lead us to mistakenly believe that such experiences seem that way.

This cognitive error theory states, that when we claim that our phenomenology of temporal binding is one as of global temporal binding, then we are expressing a false belief due to some form of cognitive error. The exact nature of that error, though, is an open question, and the aim of this section is not to settle on a general theoretical explanation of its precise nature. Trying to do so would be premature, given that the relevant data are not yet available. The goal instead is to explore how a range of different subjective inflations (each plausibly constituting a kind of cognitive error) may influence our beliefs and reports about our phenomenology of temporal binding.

Proponents of the rich view may object to this, arguing that this approach fails to take phenomenological reports at face value. However, there is an indirect way of testing whether the apparent phenomenology of richly temporally bound features in our minimally attended or unattended visual periphery is due to genuine phenomenal overflow or instead to subjective inflation. Specifically, we can test whether intervening on the factors that plausibly drive subjective

---

[36] Akin to the type described in Miller et al. (2020)

inflation reduces reports of phenomenological richness without intervening on people's discriminatory representational capacities. We can formulate another hypothesis to support H3:

> ***H4 (Subjective inflation intervention):*** People who subjectively inflate their phenomenology of experiences of temporal binding will describe their experience of temporal binding as less rich when presented with cases of temporal binding that minimise or eliminate the factors that drive subjective inflation. Moreover, such interventions will not affect people's discriminatory representational capacities related to temporal binding.

H4 makes an empirically testable claim about the existence, robustness, and malleability of the proposed subjective inflations. If H4 is true, then intervening on the factors that drive subjective inflation should selectively intervene upon people's phenomenological reports of temporal binding. This would provide strong indirect evidence for the claim that what appears as a rich phenomenology of temporal binding in our minimally attended and peripheral perceptual field is instead the result of systematic tendencies to engage in subjective inflation. If H3 and H4 are correct, then this undercuts the justification for positing merely phenomenal experiences of temporal binding that do not depend on attention.

Let us now turn to spelling out H3 and H4 by presenting the four subjective inflations in detail, together with ways in which targeted interventions on each of these may selectively affect our phenomenological reports.

### *3.3.1 Subjective Inflation from a Lack of Experiential Grasp on Temporal Binding*

The first subjective inflation I propose is due to a lack of experiential familiarity with, or sensitivity to, cases in which temporal binding fails. We rarely notice when we fail to temporally bind sensory features and thus fail to register many situations in which our experiences do not even appear temporally bound. We do not notice these failures for at least two reasons.

First, we rarely have the chance to compare our experiences to what is actually going on. This is because the world continually changes. While we may have a perception of some features as temporally bound, the environment continues to evolve, making it impossible to subjectively inspect or revisit what we experienced as temporally bound at a given moment. Second, we are rarely explicitly taking note of our temporal binding phenomenology. As a result, we may simply assume that our entire perceptual field is temporally bound, because it appears to be so in cases where we attend to it. Taken together, these two factors lead to a tendency to default to the assumption that our experiences are globally temporally bound.

We thus tend not to have an easy way to scrutinise our discriminatory representational capacities related to temporal binding. Contrast this with cases in which we are aware that we are perceiving an illusory representation of something familiar, or cases in which it is explicitly revealed to us what the illusion is and how it arises, for example, in the Müller-Lyer illusion. In those cases, we can compare our experience with how the world actually is and thereby notice our misperceptions.

I suspect that when people watch something like the colour–motion asynchrony illusion for the first time (see Animation 7), without any instruction about what to attend to or report, they would simply experience it as a normal scene. They would see a pattern of circles that change colour and change direction from left to right. However, only after being asked to make a concrete pairing between the transients produced by the two features (colour and motion direction) do they realise that this is an almost impossible task (see Animation 7 and try to pair which colour goes with which direction).

If people begin to take note of how temporal binding can fail, they may become less confident about the richness of their experiences of temporal binding. They may even start to report their experience of temporal binding in more hedged or indeterminate ways, such as: 'The colour and motion-direction seem to change together, but I would not be able to tell whether I experienced them as synchronous,' or 'The colour and motion-direction do not seem to change completely synchronously, but I am not sure that I experienced one as clearly occurring before the other,' or 'I cannot tell whether I experienced the colour and motion-direction changes as occurring together or not.'

In such cases, there would be a failure to represent the relevant features as distinctly temporally bound, even though they might still appear temporally related in a weaker or indeterminate sense. The claim is that to better understand whether one experiences sensory features as temporally bound, one must first understand what it is for temporal binding to fail (or at least to appear to fail). Once we become aware of the many situations in which features do not appear temporally bound, the apparent certainty or confidence of our phenomenological reports may be undermined, thereby deflating such descriptions.

Moreover, appropriate deflations may require explicit training, since we are not ordinarily attentive to failures of temporal binding. Importantly, That is, taking note of when our experiences are not temporally bound is not equivalent to the way in which learning a new concept enables the acquisition of a new perceptual or representational skill.

### *3.3.2 Subjective Inflation of Unity Due to Inferences Across Shifts in Attention*

A second kind of subjective inflation may stem from inferences about how things would look if we were to shift our attention. This form of subjective inflation is akin to a kind of refrigerator light illusion (Block, 2007; O'Regan & Noë, 2001; Schwitzgebel, 2009; Thomas, 1999). You might once have thought that the refrigerator light in your fridge was always on (even when the door was closed) because it is always on when you look. In much the same way, one might think that everything in one's unattended perceptual field is temporally bound because everything one attends to is experienced as temporally bound.

However, it is possible that one mistakenly conflates the *potential* for perceiving some sensory features as temporally bound (once attention is redirected) with *actually* having all sensory features in one's perceptual field represented or experienced as temporally bound. The belief that one's entire perceptual field is always temporally bound may thus result from an introspective error that conflates potential phenomenology with actual phenomenology.

The idea is that when we are asked about a particular location in the visual field, we rapidly move our attention (and often our gaze) to that location. In doing so, we fail to realise that we may not have had a consciously available, temporally bound representation of the most recent sensory information at that location prior to the shift of attention. Instead, before the shift, we may have been relying on information derived from our perception of ensemble statistics, together with predictive inferences drawn from that information, about our minimally attended experiences.

As such, multiple processes may be involved in generating this form of subjective inflation. First, we have perceptual experiences of ensemble statistics in our minimally attended perceptual field. This allows us to perceive and experience generic, low-resolution statistical features of minimally attended parts of our perceptual field (see §3.1). Second, when we shift our attention to a previously minimally attended aspect of our perceptual field, we use such predictive perceptual inferences to construct rapid perceptual interpretations of how objects are temporally bound in the newly attended region, treating ensemble statistics as a coarse perceptual model for these predictions. Third, our conscious experiences may be subject to retrospective redrafting. That is 'stimuli processed after' or 'representational content attended to' after an initial perception may interfere with or overwrite, our initial conscious perceptual interpretation (see Chapter 3 §2.1 and §4).

This combination of processes may have the effect that, after shifting our attention, the newly attended information retrospectively alters how our immediately preceding experience seems to us. As a result, the newly attended features may seem as if they were temporally bound all along. Any incongruencies or failures of temporal binding that may have been present prior to the attentional shift are thereby eliminated or rendered subjectively inaccessible, making what was not temporally bound seem temporally bound. Given the close relationship between temporal binding and attention, shifting attention may thus also shift what appears to be temporally bound.

Accordingly, in the absence of any independent ability to compare our previous experiences with our current ones, there is little reason for us to suspect that our experiences of previously minimally attended features were not always temporally bound (regardless of whether this is in fact the case). In this way, it may seem to us that everything in our perceptual field is temporally bound because temporal binding involves predictive and retrodictive processes that draft and redraft our perceptual interpretations as attention shifts (see Chapter 3 for an account of how this may occur). If this is right, then sensory features in our peripheral, minimally attended perceptual field were not temporally bound all along, even though they may vividly seem to be. This gives rise to a systematic tendency to subjectively inflate the richness of our experience of temporal binding in peripheral and minimally attended parts of our perceptual field.

#### *3.3.3 Subjective Inflation from Central and Attentional Focus*

A third kind of subjective inflation may stem from expectations based on attended and/or centrally represented perceptual content (Knotts et al., 2019). This proposal is supported by recent reviews showing how the rich detail available in attended or central vision can strongly alter the apparent experience of unattended or peripheral vision, a phenomenon known as the uniformity illusion (Otten et al., 2017; Suárez-Pinilla et al., 2018). This illusion demonstrates how our perceptual experience of peripheral stimuli changes so as to more closely match our experience of central stimuli, thereby presenting us with an apparently uniform perceptual field for features such as shape, orientation, luminance, pattern, identity, and, importantly, motion (see Figure 3 for a uniformity illusion of luminance and colour).

The apparent richness of our perceptions of otherwise sparse or indeterminate features in the periphery or minimally attended perceptual field can thus be strongly influenced by the richer and more readily available contents of attended or central perception. For example, our attended and central perception of a specific motion pattern can influence our apparent peripheral visual experience of a distinct motion pattern, to the point that the peripheral pattern seems phenomenologically identical to the central one. In this way, attended and central perception can give rise to apparently rich but subjectively inflated phenomenology of peripheral vision.

**Figure 3:** A colour luminance version of the uniformity illusion. When the rectangular figure fills your entire visual field and you fix your gaze at the centre for at least 20 seconds you should start feeling the strength of the uniformity illusion (adapted from Otten et al., 2017). You should see the colour in the centre bleed into the periphery, so you at some point end up with an entirely uniform colour.

A similar effect may plausibly hold for temporal binding, given that we are subject to a motion-based version of the uniformity illusion. If attended or central sensory features are perceived as temporally bound, and sensory features in our minimally attended or peripheral perceptual field change in a similar yet not strictly identical way, then we may take ourselves to have a matching experience of temporal binding in those minimally attended or peripheral regions. That is, the perception of temporal binding in our central visual field may uniformly influence the apparent phenomenology of temporal binding in the periphery.

There is already empirical evidence showing that the timing of attended sensory features affects our perception of the timing of other, less salient features. Consider the case of temporal ventriloquism (discussed in Chapters 2, 3, and 4) where attending to a specific feature (a sound) can cause us to perceive another, less salient feature (a flash of light) as occurring simultaneously, or nearly simultaneously, with the sound, regardless of whether the flash objectively occurs before or after it (Hartcher-O’Brien & Alais, 2011).

Similarly, I predict that the temporal binding of central and attended sensory features can ‘bleed into’ our peripheral vision, strongly shaping what we report about the temporal relations of peripheral sensory features and ensemble statistical information (see Animation 8). I refer to this effect as the temporal binding uniformity illusion.

In Animation 8, the centre of the display contains a standard 30-frame colour–motion pattern: 15 frames of one colour (red) moving right, followed by 15 frames of another colour (green) moving left, presented at 120 FPS. When the video fills one's visual field, the peripheral pattern becomes phenomenologically indistinguishable from the central pattern, despite the fact that the two patterns are objectively out of sync—changing colour and motion direction at different times and moving at different speeds. This can be seen by comparing Animation 8 with the control case in Animation 9, where there is only a single, genuinely uniform pattern.

If participants report that the minimally attended colour–motion pattern is identical to the central colour–motion pattern, this will provide evidence that they are subjectively inflating their phenomenology of temporal binding on the basis of attended and centrally represented perceptual content, rather than on the basis of temporally bound representations available in our peripheral perceptions.

### *3.3.4 Subjective Inflation from Prior Knowledge*

A fourth kind of subjective inflation comes from expectations grounded in prior knowledge. This form of subjective inflation resembles the inflation driven by prediction and central perception, but it is more narrowly concerned with inferences based on prior knowledge, rather than with predictive inferences across shifts of attention. This kind of subjective inflation is supported by a series of studies reviewed by Knotts et al. (2019).

In one of the reviewed studies by Li et al. (2002) shows how the subjective inflation of an object's colour depends on expectations about that object's typical colour. De Gardelle et al. (2009) and Balas & Sinha (2007) show how nonsensical characters located in minimally attended peripheral vision give rise to apparent experiences of standard alphanumeric characters when participants are engaged in a reading task. Similar effects have also been observed for peripheral motion perception (Zhang et al., 2013) and object recognition (Wijntjes & Rosenholtz, 2018).

Accordingly, it may be that expectancy (shaped by prior knowledge about how features typically temporally relate) strongly influences how we describe the apparent phenomenology of how sensory features in our minimally attended perceptions are temporally related. If the influence of such expectations is sufficiently strong, this may lead us to describe minimally attended features as appearing to us to be temporally bound, even in the absence of attention-dependent representational support for such binding.

It is evident from a range of studies that predictions strongly influence how sensory features are temporally bound at the level of perceptual representation (see Chapter 3, §2.2). These same predictive mechanisms may therefore also influence how we characterise our experience in the peripheral and minimally attended perceptual field. That is, we have strong expectations about the temporal relations of sensory features, and if peripheral ensemble statistics are shaped by such expectations, then our perceptual field may appear globally temporally bound, despite the underlying experience being limited to averaged, predicted statistical temporal relations. This, in turn, may lead us to infer and report that our minimally attended peripheral experiences are richly temporally bound in much the same way as our central, attended experiences. Hence, subjectively inflating our phenomenology of temporal binding in our peripheral visual field.

### *3.3.5 Upshot of Subjective Inflation Explanations*

Taken together, these different forms of subjective inflation may explain why we describe our experiences as globally temporally bound, even if the underlying perceptual representations are constrained by attentional capacities. Each of these explanations can be tested individually to determine how, and to what extent, the corresponding subjective inflation affects reported phenomenology of temporal binding. Moreover, it is plausible that these different forms of subjective inflation share a common underlying factor, namely, *expectancy*.

I have argued that the subjective inflation of experiences of temporal binding under conditions of minimal attention is influenced by:

(i) The expectation that everything is temporally bound, leading to a failure to notice when experiences are not temporally bound.

(ii) Pre- and post-perceptual inferential expectations about how sensory features in the peripheral field would be experienced as temporally bound when, and after, attention is shifted.

(iii) Inferential expectations grounded in attended or central perception, whereby experiences of temporal binding in the attended central visual field shape apparent experiences of temporal binding in the periphery.

(iv) Inferential expectations based on prior knowledge about how sensory features typically tend to be experienced as temporally bound.

The fact that expectancy plays a role in all four forms of subjective inflation suggests that future empirical tests should explicitly target the general relationship between expectancy and subjective inflation. There are two important advantages to focusing on expectancy. First, expectancy is functionally separable from attention. Second, expectancy is an easily manipulable variable. Expectancy depends primarily on the predictability of target stimuli, whereas attention depends on the mental selection of target stimuli. It is therefore possible to manipulate expectation by altering predictability without intervening in the degree of attention directed at the target stimuli, and vice versa (Summerfield & Egner, 2009; Zhou et al., 2022; Zivony & Eimer, 2024).

If subjective inflation of our phenomenology of temporal binding critically depends on expectancy, then the inflation effect should be stronger when subjects attend to locations where target stimuli are expected (in one or more of the four ways described above) than when stimuli occur in less expected locations or less expected ways.

At this point, we might imagine a scenario in which participants persistently maintain that their peripheral and minimally attended perceptions are experienced as richly temporally bound, despite evidence that they subjectively inflate their phenomenology in the ways outlined above, and despite recognising that they perform poorly on peripheral temporal binding tasks. In such cases, we may observe a divergence between how participants adjust their confidence in what they report and how they describe the richness of their phenomenology. Participants may for pragmatic reasons reduce their confidence in their reports (otherwise they would knowingly lose money by betting on their apparent phenomenology) but do so without revising their description of their phenomenology.

The experience of temporal binding may thus be phenomenologically ‘sticky’, making it difficult to relinquish the sense that one’s entire perceptual field is richly temporally bound, even when one cannot successfully discriminate temporal relations among minimally attended features. Some may insist that they experience more than mere gist-level representations, and that their experiences cannot be explained by selective central attention combined with peripheral representational limits. However, such a position cannot be justified unless one can provide empirical evidence that H2 is false (i.e., that there is no discrepancy between capacity and phenomenology) and that H3 is false (i.e., that subjective inflation does not account for the apparent richness of experience). Absent such evidence, the claim rests solely on intuition and stands in tension with empirically supported claims.

## 4. Conclusion

The motivation behind this chapter is that we need new kinds of empirical data to answer the question of how our phenomenology of temporal binding depends on attention. Empirically investigating the phenomenology of people's experiences of temporal binding is a largely novel undertaking, and new data in this domain can teach us something important about how consciousness employs temporal binding as a way of structuring the world into coherent perceptual wholes, thereby providing a means of putting competing theories of temporal binding to the test. I have argued that there are principled ways of getting hold of such data, and I have outlined what different theoretical views should lead us to expect.

This chapter develops both a methodological and a theoretical framework for interpreting and generating new experimental paradigms aimed at testing how experiences of temporal binding depend on attention. Future empirical work building on this framework will plausibly proceed in several stages.

First, one would need to develop the temporal binding tasks suggested in §3.1 to include systematic manipulations of attention, thereby allowing for comparisons of people's experiences across a range of attentional conditions. Second, one must find a way to implement the report protocols suggested in §3.1 within controlled experimental settings, in order to gather detailed and structured reports about people's experiences of temporal binding. Third, to confirm or disconfirm the central hypothesis introduced in §3.2, one would need to establish a protocol for systematically comparing phenomenological reports and task performance, specifically examining whether and how these are differently affected by attentional manipulation. Finally, to assess the extent to which people's reports of their experiences of temporal binding are subject to subjective inflation, one must develop studies along the lines proposed in §3.3. Such studies would help determine whether any observed discrepancies between phenomenological reports and discriminatory representational capacities are best explained in terms of phenomenal overflow or, instead, in terms of attentional constraints combined with systematic cognitive error.

Taken together, the framework developed in this chapter outlines a principled way of empirically investigating how attention influences the content and phenomenology of our experiences of temporal binding, clarifying different theoretical views and how they can be tested.

# Conclusion

I have argued that a theory of temporal binding should address three explanatory problems. It should: (i) account for the psychophysical data on how a wide variety of effects influence temporal binding, (ii) address for the problem of neural delay and the problem of desynchronisation, and (iii) explicate the relation between the mechanisms underlying temporal binding and the content and phenomenology of our experiences of temporal binding.

I argued, in Chapter 2, that the brain time theory was unable to address problems (i) and (ii), because its definition of temporal binding in terms of synchrony of neural processing times lacked the tools to account for the flexibility and adaptability of temporal binding, which is not due to differences in neural processing times. As such, I concluded that temporal binding is not fully determined by differential latencies in neural processing.

In Chapter 3 I set out a novel hybrid theory of temporal binding as an alternative to the brain time theory. This theory proposes two novel hypotheses. The first states that our perceptual system relies on a hybrid system of retrodictive processing and predictive processing of sensory features to engage in temporal binding. The second hypothesis states that attentional modulation determines which processed sensory features become part of a temporally bound perception. With these in place, the hybrid theory is well placed to address Problems (i) and (ii). First, the hybrid theory can account for (i). By flexibly deploying retrodictive and predictive processing depending on how attention is probed, temporal binding becomes sensitive to context-dependent factors such as goals, intentions, causal regularities, expectations, salience of stimuli, and our actions. Second, it can account for (ii). Retrodictive redrafting allows for the perceptual compensation of desynchronisation in sensory processing, and predictive drafting allows the perceptual compensation for neural delays in processing times. A few consequences follow from this view:

- Temporal binding is inherently *flexible* such that it can occur in variable ways and *adaptive* in that it is responsive to changing internal and external demands.

- The onsets/offsets of conscious experiences are temporally indeterminate, i.e., there are no temporally determinate ‘now’ tokens of our reported conscious experience. The content of any alleged ‘now’ token is spread across an amalgam of predictive perceptual drafts that are continuously retrodictively redrafted.
- It is likely that temporally bound perceptions are necessarily conscious. If contents are attentionally modulated in the relevant way to be temporally bound, they will also likely be consciously represented.
- Our experiences (both content and phenomenology) of temporal binding are limited by attentional capacities.

Chapters 4 and 5 show how this theory also provides an answer to Problem (iii). In Chapter 4, I spell out how the hybrid theory explains in what ways we experience time as perceptually unified, albeit in a deflated sense, and how this experience emerges from the widely fragmented and dissociable mechanisms that underlie time perception. More specifically, I argued that the following three claims about our perceived unity of time are entailed by the hybrid theory:

- Our phenomenology of what we perceive as a unified ‘now’ is sparse, malleable, and constrained by attention.
- The phenomenology that our experience is organized within a coherent determinate unified timeline is contingent and at best a representation of short, coherent segments maintained in working memory, that are overwritten as our attention shifts to focus on different gestalts.
- The comparability of temporal properties is structured by the coherence imposed by selective attention and predictive and retrodictive processing.

In Chapter 5, I present and defend what I call the sparse temporal binding view, which explains how attention and our phenomenology and representational content of temporal binding relate. In short, the view states that we enjoy rich perceptual experiences of temporal binding of sensory features within the central and selectively attended perceptual field, and at best sparse gist perceptual information of temporal relations of sensory features in our peripheral and minimally attended perceptual field. The chapter proposes a framework for empirically testing how our experience of temporal binding depends on attention. In doing so I set out three hypotheses that followed from the hybrid theory:

- The richness of perceptual experiences of temporal binding is directly affected by the degree to which we can attend to the relevant features.
- There should be no discrepancy between how people's phenomenology and discriminatory representational capacities related to temporal binding depend on attention.
- If there are any discrepancies between people's phenomenology and discriminatory representational capacities, this is due to systematic subjective inflation of temporal binding.

**Avenues for Future Work**

The hybrid theory proposes several necessary functional mechanisms which generate several clear avenues for future work. I have set out some of the most pressing ones below:

(i) *Targeted Investigation of the mechanisms proposed by the hybrid theory:* Empirical studies should probe the extent and conditions under which retrodictive and predictive processing shapes temporal binding. This is an area that is rife with conceptual, theoretical, and empirical challenges. Many perceptual effects, such as the flash-lag effect, have multiple competing explanations involving predictive, retrodictive, or latency-based mechanisms (see Eagleman & Sejnowski, 2000, 2007; Hogendoorn, 2020; Holcombe & Corbett, 2023; Nijhawan, 1994; Patel et al., 2000). Much more empirical work is needed to test the discrete effects that are likely due to retrodictive and predictive processing.

(ii) *Explore the Relation Between Consciousness and Temporal Binding:* While the hybrid theory predicts that temporal binding is likely to be a necessarily conscious phenomenon, such that we do not temporally bind features outside consciousness, this claim is not substantiated by empirical research. In fact, there is no such research. As discussed in Chapter 3, §5.2 blindsight at best shows evidence of unconscious spatial binding, yet not of temporal binding. So new theoretical and empirical research is needed to better understand whether or not unconscious temporal binding is possible.

(iii) *Empirical Testing of Attention's Role in Temporal Binding:* Systematic manipulations of attention in temporal binding tasks are needed to test the theory's central claim that attentional modulation determines the contents of temporally bound perception. Paradigms such as those proposed in Chapter 5 should be used to investigate the relationship between attention and (a) discriminatory representational capacities of temporal binding and (b) phenomenology of temporal binding.

(iv) *Investigating Subjective Inflation of Temporal Binding:* A core hypothesis of the hybrid theory is that phenomenological reports may be subjectively inflated under conditions of minimal attention. Future experiments should test whether different kinds of expectancy (as those proposed in Chapter 5, §3.3) drive subjective inflations by seeing if eliminating these driving forces reduces the reported richness of phenomenology without affecting people's discriminatory representational capacities.

(v) *Extending the Framework from Chapter 4:* Further work is needed on how the dissociable neural mechanisms underlying time perception contribute to the different forms of temporal unity described in Chapter 4. This includes collecting detailed phenomenological data under conditions that selectively intervene on specific mechanisms of temporal processing.

(vi) *Application to the philosophical literature on temporal consciousness:* The hybrid theory provides a new angle on philosophical debates on the temporal structure of consciousness. It challenges views that assume that the temporal structure of conscious perception mirrors the structure of neural processing by emphasizing that temporal binding emerges from higher-order temporal organisation. As such the theory is in outright conflict with extensionalist theories that embrace this kind of mirroring (Dainton, 2000; Phillips, 2014; Piper, 2019; Rashbrook, 2013). It also sits awkwardly with snapshot theories that suggest that we have instant but dynamic snapshot-like experience (Arstila, 2018; Le Poidevin, 2007; Prosser, 2016). While not denying that we can encode dynamic content at an instant, the hybrid theory rejects its assumption that the temporal organisation of perception follows from the temporal organisation of the processing of sensory features. The hybrid theory can therefore only be interpreted as a retentionalist theory, some kind of anti-realist theory, or a different theory altogether. Further work is needed to draw these connections.

Taken together, this thesis provides a unified conceptual, empirical, and methodological framework for understanding temporal binding. It shows how our perceptual system flexibly integrates sensory features into temporally bound experiences through the interaction of predictive and retrodictive processing, with attention playing a central role in selecting which features enter conscious awareness. This hybrid theory accounts for both the flexibility and constraints of temporal binding, explains the apparent unity of time despite fragmented underlying mechanisms, and deflates our phenomenology, explaining why we describe our perception of time as richer and more unified than it is.

# References

[1] Abramov, I., Gordon, J., & Chan, H. (1991). Color appearance in the peripheral retina: Effects of stimulus size. *Journal of the Optical Society of America. A, Optics and Image Science*, *8*(2), 404–414. https://doi.org/10.1364/josaa.8.000404

[2] Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure & Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

[3] 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

[4] 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

[5] Akyürek, E. G. (2025). Temporal integration as an adaptive process in visual perception, attention, and working memory. *Neuroscience & Biobehavioral Reviews*, *170*, 106041. https://doi.org/10.1016/j.neubiorev.2025.106041

[6] Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, *14*(3), 257–262. (14761661). https://doi.org/10.1016/j.cub.2004.01.029

[7] 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

[8] Alexander, H. G. (1956). The Leibniz-Clarke Correspondence. Philosophy 32 (123):365-366.

[9] Allman, M. J., & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain: A Journal of Neurology*, *135*(Pt 3), 656–677. https://doi.org/10.1093/brain/awr210

[10] Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the internal clock: First- and second-order principles of subjective time. *Annual Review of Psychology*, *65*, 743–771. https://doi.org/10.1146/annurev-psych-010213-115117

[11] Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. Psychological Science, 19(4), 392–398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

[12] Anderson, B. (2011). There is no Such Thing as Attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00246

[13] Anderson, P. W., & Zahorik, P. (2014). Auditory/visual distance estimation: Accuracy and variability. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01097

[14] (Anonymous) E. Robert Kelly; aka E.R. Clay]. (1882). *The Alternative: A Study in Psychology*. London: Macmillan and Co.

[15] Aristotle. (Physics, Volume I: Books 1-4). *Physics,: I* (Translated by P. H. Wicksteed, F. M. Cornford. Loeb Classical Library 228. Cambridge, MA: Harvard University Press, 1957.).

[16] Arnold, D. H. (2010). Relative timing and perceptual asynchrony. In B. Khurana & R. Nijhawan (Eds.), *Space and Time in Perception and Action* (pp. 254–277). Cambridge University Press. Cambridge Core. https://doi.org/10.1017/CBO9780511750540.016

[17] 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

[18] Arrighi, R., Alais, D., & Burr, D. (2005). Perceived timing of first- and second-order changes in vision and hearing. *Experimental Brain Research*, *166*(3–4), 445–454. https://doi.org/10.1007/s00221-005-2384-9

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

[20] 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

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

[22] 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

[23] Arstila, V. (2018). Temporal Experiences without the Specious Present. *Australasian Journal of Philosophy*, *96*(2), 287–302. https://doi.org/10.1080/00048402.2017.1337211

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

[25] Arstila, V. (2023). Explanation in theories of the specious present. *Philosophical Psychology*, *0*(0), 1–24. https://doi.org/10.1080/09515089.2023.2241501

[26] Aru, J., & Bachmann, T. (2013). Phenomenal awareness can emerge without attention. *Frontiers in Human Neuroscience*, *7*, 891. https://doi.org/10.3389/fnhum.2013.00891

[27] Augustine, S. (401 C.E.). *The Confessions of Saint Augustine*. Project Gutenberg, Translation by E. B. Pusey (Edward Bouverie). https://www.gutenberg.org/files/3296/3296-h/3296-h.htm

[28] Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

[29] Azzopardi, P., & Cowey, A. (2001). Motion discrimination in cortically blind patients. Brain: A Journal of Neurology, 124(Pt 1), 30–46. https://doi.org/10.1093/brain/124.1.30

[30] 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

[31] Baars, B., Franklin, S., & Ramsøy, T. (2013). Global Workspace Dynamics: Cortical “Binding and Propagation” Enables Conscious Contents. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00200

[32] Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

[33] Baars, B. J. (1993). How does a serial, integrated and very limited stream of consciousness emerge from a nervous system that is mostly unconscious, distributed, parallel and of enormous capacity? *Ciba Foundation Symposium*, *174*, 282–290; discussion 291-303. https://doi.org/10.1002/9780470514412.ch14

[34] Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends Cogn Sci*, *6*(1), 47–52. (11849615). https://doi.org/10.1016/s1364-6613(00)01819-2

[35] 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

[36] Badcock, P. B., Friston, K. J., & Ramstead, M. J. D. (2019). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of Life Reviews, Physics of Mind*, *31*, 104–121. https://doi.org/10.1016/j.plrev.2018.10.002

[37] Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *49A*(1), 5–28. https://doi.org/10.1080/027249896392784

[38] Baddeley, A. D. (1993). Verbal and visual subsystems of working memory. *Current Biology: CB*, *3*(8), 563–565. https://doi.org/10.1016/0960-9822(93)90059-w

[39] 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

[40] Balas, B., & Sinha, P. (2007). "Filling-in" colour in natural scenes. *Visual Cognition*, *15*(7), 765–778. https://doi.org/10.1080/13506280701295453

[41] Bayne, T. (2010). *The Unity of Consciousness*. Oxford University Press UK.

[42] Bayne, T. J., & Chalmers, D. J. (2003). What is the Unity of Consciousness? In A. Cleeremans (Ed.), *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford University Press.

[43] 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

[44] Bechlivanidis, C., & Lagnado, D. (2009). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition*, *146*, 58–66. https://doi.org/10.1016/j.cognition.2015.09.001

[45] 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. (12972391). https://doi.org/10.1016/s0042-6989(03)00436-x

[46] Behm, D. G., & Carter, T. B. (2020). Effect of Exercise-Related Factors on the Perception of Time. *Frontiers in Physiology*, *11*. https://doi.org/10.3389/fphys.2020.00770

[47] 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

[48] 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

[49] 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

[50] Block, Ned. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *The Behavioral and Brain Sciences*, *30*(5–6), 481–499; discussion 499-548. https://doi.org/10.1017/S0140525X07002786

[51] Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. https://doi.org/10.1016/j.tics.2011.11.001

[52] Block, N. (2014). Seeing-As in the Light of Vision Science. *Philosophy and Phenomenological Research*, *89*(3), 560–572. https://doi.org/10.1111/phpr.12135

[53] Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends Cogn Sci*, *23*(12), 1003–1013. (31676213). https://doi.org/10.1016/j.tics.2019.10.001

[54] 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

[55] 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

[56] 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

[57] Boethius, A. (2000). *The Consolation of Philosophy: Revised Edition*. Translated by Victor Watts, Penguin Publishing Group.

[58] Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624

[59] Bowman, H., Filetti, M., Wyble, B., & Olivers, C. (2013). Attention is more than prediction precision. *The Behavioral and Brain Sciences*, *36*(3), 206–208. https://doi.org/10.1017/S0140525X12002324

[60] Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology*, *2*(1), 1–3. https://doi.org/10.4161/cib.2.1.7297

[61] Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. Psychological Review, 120(1), 85–109. https://doi.org/10.1037/a0030779

[62] Brigard, F. de, & Prinz, J. (2010). Attention and Consciousness. *Wiley Interdisciplinary Reviews*, *1*(1), 51–59.

[63] Broad, C. D. (1923). *Scientific Thought*. Routledge and Kegan Paul.

[64] Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: “Cost free” color phenomenality outside focal attention. *Psychological Science*, *25*(7), 1394–1403. https://doi.org/10.1177/0956797614532656

[65] Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. https://doi.org/10.1016/j.tics.2019.06.009

[66] Brown, T. (1857). *Philosophy of the Human Mind*. Edinburgh: Adam & Charles Black.

[67] Buchsbaum, B. R. (2013). The role of consciousness in the phonological loop: Hidden in plain sight. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00496

[68] Buehner, M. J. (2010). Temporal binding. In A. C. Nobre & J. T. Coull (Eds.), Attention and Time (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199563456.003.0015

[69] Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, The Brain in Health and Disease - from Molecules to Man. Swiss National Foundation Symposium NRP 38*, *36*(2), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

[70] Burgoyne, A. P., & Engle, R. W. (2020). Attention Control: A Cornerstone of Higher-Order Cognition. *Current Directions in Psychological Science*, *29*(6), 624–630. https://doi.org/10.1177/0963721420969371

[71] Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nat Neurosci*, *10*(4), 423–425. (17369824). https://doi.org/10.1038/nn1874

[72] Busch, N. A. (2013). The fate of object memory traces under change detection and change blindness. *Brain Research*, *1520*, 107–115. https://doi.org/10.1016/j.brainres.2013.05.014

[73] Callender, C. (2017). *What Makes Time Special?* Oxford University Press.

[74] Cha, O., & Chong, S. C. (2018). Perceived Average Orientation Reflects Effective Gist of the Surface. Psychological Science, 29(3), 319–327. https://doi.org/10.1177/0956797617735533

[75] Chakravarthi, R., & Vanrullen, R. (2012). Conscious updating is a rhythmic process. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), 10599–10604. https://doi.org/10.1073/pnas.1121622109

[76] Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

[77] Chalmers, D. J. (2006). Perception and the Fall From Eden. In T. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 49–125). Oxford University Press.

[78] Chalmers, D. J. (2010). *The Character of Consciousness*. Oxford University Press.

[79] Chappell, M., & Hinchy, J. (2014). Turning the corner with the flash–lag illusion. *Vision Research*, *96*, 39–44. https://doi.org/10.1016/j.visres.2013.12.002

[80] Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, *130*(3), 225–234. https://doi.org/10.1016/j.actpsy.2008.12.008

[81] Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception & Psychophysics*, *75*(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

[82] Chen, W.-Y., Howe, P. D., & Holcombe, A. O. (2013). Resource demands of object tracking and differential allocation of the resource. *Attention, Perception, & Psychophysics*, *75*(4), 710–725. https://doi.org/10.3758/s13414-013-0425-1

[83] 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

[84] Chisholm, R. M. (1981). Brentano's Analysis of the Consciousness of Time. *Midwest Studies in Philosophy*, *6*(1), 3–16.

[85] Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/BF03195009

[86] Chota, S., & VanRullen, R. (2019). Visual Entrainment at 10 Hz Causes Periodic Modulation of the Flash Lag Illusion. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00232

[87] Chuard, P. (2017). The Snapshot Conception of Temporal Experiences. In *The Routledge Handbook of Philosophy of Temporal Experience*. Routledge.

[88] Church, R. M. (1984). Properties of the internal clock. *Annals of the New York Academy of Sciences*, *423*, 566–582. https://doi.org/10.1111/j.1749-6632.1984.tb23459.x

[89] Clark, Andy. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

[90] Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. OUP USA.

[91] Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, *17*(3), 521–534. https://doi.org/10.1007/s11097-017-9525-z

[92] Clark, A. (2019). Consciousness as Generative Entanglement. *The Journal of Philosophy*, *116*(12), 645–662. https://doi.org/10.5840/jphil20191161241

[93] Clark, A., Friston, K., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. *Journal of Consciousness Studies*, *26*(9–10), 19–33. Scopus.

[94] Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science*, *22*(9), 1165–1172. https://doi.org/10.1177/0956797611419168

[95] Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, *16*(8), 411–417. https://doi.org/10.1016/j.tics.2012.06.013

[96] Cohen, M. A., & Chun, M. M. (2017). Studying Consciousness Through Inattentional Blindness, Change Blindness, and the Attentional Blink. In *The Blackwell Companion to Consciousness* (pp. 537–550). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119132363.ch38

[97] Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, *15*(8), 358–364. https://doi.org/10.1016/j.tics.2011.06.008

[98] 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

[99] Cohen, M. A., Ostrand, C., Frontero, N., & Pham, P.-N. (2021). Characterizing a snapshot of perceptual experience. *Journal of Experimental Psychology. General*, *150*(9), 1695–1709. https://doi.org/10.1037/xge0000864

[100] Cohen, M. A., & Rubenstein, J. (2020). How much color do we see in the blink of an eye? *Cognition*, *200*, 104268. https://doi.org/10.1016/j.cognition.2020.104268

[101] Colombo, M., & Wright, C. (2017). Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain and Cognition*, *112*, 3–12. https://doi.org/10.1016/j.bandc.2016.02.003

[102] Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

[103] 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

[104] 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

[105] Cowan N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences 2001; 24:87–114. doi: 10.1017/s0140525x01003922.

[106] Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention During Natural Vision Warps Semantic Representation Across the Human Brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

[107] Cunningham, D. W., Billock, V. A., & Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychol Sci*, *12*(6), 532–535. (11760144). https://doi.org/10.1111/1467-9280.d01-17

[108] Dainton, B. (2000). *Stream of Consciousness: Unity and Continuity in Conscious Experience*. Routledge.

[109] Dainton, B. (2001). *Time and Space*. Mcgill-Queen's University Press.

[110] Dainton, B. (2008). Sensing change. *Philosophical Issues*, *18*(1), 362–384.

[111] Dainton, B. (2024) Temporal Consciousness. in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta & Uri Nodelman (eds.),

[112] Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/s0042-6989(97)00133-8

[113] de Gardelle, V., Sackur, J., & Kouider, S. (2009). Perceptual illusions in brief visual presentations. *Consciousness and Cognition*, *18*(3), 569–577. https://doi.org/10.1016/j.concog.2009.03.002

[114] 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

[115] Dehaene, S. (2014). *Consciousness and the Brain—Deciphering How the Brain Codes Our Thoughts*. Penguin Group.

[116] Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn Sci*, *10*(5), 204–211. (16603406). https://doi.org/10.1016/j.tics.2006.03.007

[117] Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

[118] Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. (11164022). https://doi.org/10.1016/s0010-0277(00)00123-2

[119] 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

[120] Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia*, *46*(11), 2861–2864. https://doi.org/10.1016/j.neuropsychologia.2008.05.014

[121] 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

[122] Dennett, D. C. (1988). The Myth of Double Transduction. *Tucson*, *volume II*, 97–107.

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

[124] Dennett, D. C. (1992). Filling in versus finding out: A ubiquitous confusion in cognitive science. In H. Pick, P. Van den Broek, & D. Knill (Eds.), *[Book Chapter]*. American Psychological Association.

[125] 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.

[126] Derrien, D., Garric, C., Sergent, C., & Chokron, S. (2022). The nature of blindsight: Implications for current theories of consciousness. *Neuroscience of Consciousness*, *2022*(1), niab043. https://doi.org/10.1093/nc/niab043

[127] Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

[128] Devinsky, O., & D'Esposito, M. (2004). *Neurology of Cognitive and Behavioral Disorders*. Oxford University Press, USA.

[129] Di Lollo, V. (2018). Attention is a sterile concept; iterative reentry is a fertile substitute. *Consciousness and Cognition, Visual Experience and Guidance of Action: A Tribute to Bruce Bridgeman*, *64*, 45–49. https://doi.org/10.1016/j.concog.2018.02.005

[130] Di Lollo, V., Lowe, D. G., & Scott, J. P. (1974). Backward masking and interference with the processing of brief visual displays. *Journal of Experimental Psychology*, *103*(5), 934–940. https://doi.org/10.1037/h0037362

[131] 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

[132] Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719–721. https://doi.org/10.1068/p090719

[133] 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

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

[135] Dretske, F. (2007). What Change Blindness Teaches about Consciousness. *Philosophical Perspectives*, *21*, 215–230.

[136] 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

[137] Drissi-Daoudi, L., Öğmen, H., & Herzog, M. H. (2021). Features integrate along a motion trajectory when object integrity is preserved. *Journal of Vision*, *21*(12), 4. https://doi.org/10.1167/jov.21.12.4

[138] Droege, P. (2009). Now or never: How consciousness represents time. *Consciousness and Cognition*, *18*(1), 78–90. https://doi.org/10.1016/j.concog.2008.10.006

[139] 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

[140] Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131–136. https://doi.org/10.1016/j.conb.2008.06.002

[141] Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, *287*(5460), 2036–2038. (10720334). https://doi.org/10.1126/science.287.5460.2036

[142] Eagleman, D. M., & Sejnowski, T. J. (2007). Motion signals bias localization judgments: A unified explanation for the flash-lag, flash-drag, flash-jump, and Frohlich illusions. *Journal of Vision*, *7*(4), 3–3. https://doi.org/10.1167/7.4.3

[143] 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

[144] 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

[145] Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, *369*(6481), 525. https://doi.org/10.1038/369525a0

[146] Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, *67*(5), 279–300. https://doi.org/10.1037/h0041622

[147] Fagelson, M. A., & Champlin, C. A. (1997). Auditory filters measured at neighboring center frequencies. *The Journal of the Acoustical Society of America*, *101*(6), 3658–3665. https://doi.org/10.1121/1.418326

[148] Faivre, N., Arzi, A., Lunghi, C., & Salomon, R. (2017). Consciousness is more than meets the eye: A call for a multisensory study of subjective experience†. *Neuroscience of Consciousness*, *2017*(1), nix003. https://doi.org/10.1093/nc/nix003

[149] Ferretti, G., & Glenney, B. (Eds.). (2020). Molyneux's Question and the History of Philosophy (1st ed.). Routledge.

[150] Finke, R. A., Freyd, J. J., & Shyi, G. C. (1986). Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology. General*, *115*(2), 175–188. https://doi.org/10.1037//0096-3445.115.2.175

[151] 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

[152] Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

[153] Fontes, R., Ribeiro, J., Gupta, D. S., Machado, D., Lopes-Júnior, F., Magalhães, F., Bastos, V. H., Rocha, K., Marinho, V., Lima, G., Velasques, B., Ribeiro, P., Orsini, M., Pessoa, B., Leite, M. A.

A., & Teixeira, S. (2016). Time Perception Mechanisms at Central Nervous System. *Neurology International*, *8*(1), 5939. https://doi.org/10.4081/ni.2016.5939

[154] Fornaciai, M., & Di Luca, M. (2020). Causality Shifts the Perceived Temporal Order of Audiovisual Events. *Journal of Experimental Psychology: Human Perception and Performance*, *46*. https://doi.org/10.1037/xhp0000754

[155] Fornaciai, M., Markouli, E., & Di Luca, M. (2018). Modality-specific temporal constraints for state-dependent interval timing. *Scientific Reports*, *8*(1), 10043. https://doi.org/10.1038/s41598-018-28258-4

[156] Foster, J. (1982). *The Case for Idealism* (Issue 3, pp. 465–468). Routledge.

[157] Foster, J. (1991). *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. Routledge.

[158] Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology. Human Perception and Performance*, *33*(5), 1003–1012. https://doi.org/10.1037/0096-1523.33.5.1003

[159] Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39.

[160] Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(5), 1738–1747. https://doi.org/10.1523/JNEUROSCI.4403-13.2014

[161] Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

[162] Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

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

[164] Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

[165] Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*(1), 220–234. https://doi.org/10.1016/j.neuroimage.2006.08.035

[166] Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat Neurosci*, *7*(7), 773–778. (15195098). https://doi.org/10.1038/nn1268

[167] Fujisaki, Waka, Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), Article 7. https://doi.org/10.1038/nn1268

[168] Gallagher, S., & Zahavi, D. (2001). Primal impression and enactive perception. *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality*, 83–100.

[169] Gallagher, S., & Zahavi, D. (2021). *The Phenomenological Mind*. Third edition, Routledge.

[170] García-Pérez, M. A., & Alcalá-Quintana, R. (2015). Converging evidence that common timing processes underlie temporal-order and simultaneity judgments: A model-based analysis. *Attention, Perception, & Psychophysics*, *77*(5), 1750–1766. https://doi.org/10.3758/s13414-015-0869-6

[171] 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

[172] Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, *423*, 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

[173] Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: Advances and challenges. *Current Opinion in Neurobiology*, *7*, 170–184.

[174] Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

[175] Goodale, M. A., Westwood, D. A., & Milner, A. D. (2004). Two distinct modes of control for object-directed action. *Prog Brain Res*, *144*, 131–144. (14650845). https://doi.org/10.1016/s0079-6123(03)14409-3

[176] Gordon, J., & Abramov, I. (1977). Color vision in the peripheral retina. II. Hue and saturation. *Journal of the Optical Society of America*, *67*(2), 202–207. https://doi.org/10.1364/josa.67.000202 66.

[177] Greene MR, Oliva A. Recognition of natural scenes from global properties: Seeing the forest without representing the trees. Cognit. Psychol 2009;58:137–176. [PubMed: 18762289]

[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

[179] Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A. F., & Scholte, H. S. (2013b). From Image Statistics to Scene Gist: Evoked Neural Activity Reveals Transition from Low-Level Natural Image Structure to Scene Category. Journal of Neuroscience, 33(48), 18814–18824. https://doi.org/10.1523/JNEUROSCI.3128-13.2013

[180] 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. (16135885). https://doi.org/10.1088/1741-2560/2/3/s05

[181] Haddad, R., Khan, R., Takahashi, Y. K., Mori, K., Harel, D., & Sobel, N. (2008). A metric for odorant comparison. *Nature Methods*, *5*(5), 425–429. https://doi.org/10.1038/nmeth.1197

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

[183] Hamilton, W. (1860). *Lectures on metaphysics and logic*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

[184] Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision*, *9*(4), 26.1-12. https://doi.org/10.1167/9.4.26

[185] Harrar, V., & Harris, L. R. (2005). Simultaneity constancy: Detecting events with touch and vision. *Experimental Brain Research*, *166*(3), 465–473. https://doi.org/10.1007/s00221-005-2386-7

[186] Harrison, L. M., Bestmann, S., Rosa, M. J., Penny, W., & Green, G. G. R. (2011). Time Scales of Representation in the Human Brain: Weighing Past Information to Predict Future Events. *Frontiers in Human Neuroscience*, *5*, 37. https://doi.org/10.3389/fnhum.2011.00037

[187] Hartcher-O'Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(5), 1383–1395. https://doi.org/10.1037/a0024234

[188] Hashimoto, M., & Kumashiro, M. (2004). [Intermodal timing cues for audio-visual speech recognition]. *Journal of UOEH*, *26*(2), 215–225. https://doi.org/10.7888/juoeh.26.215

[189] Hassenstein, B., & Reichardt, W. (1956). Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Z. Naturforsch.*, *11b*, 513–524. https://doi.org/10.1515/znb-1956-9-1004

[190] Haun, A. M., Tononi, G., Koch, C., & Tsuchiya, N. (2017). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness*, *2017*(1), niw023. https://doi.org/10.1093/nc/niw023

[191] Hawkins, B., Evans, D., Preston, A., Westmoreland, K., Mims, C. E., Lolo, K., Rosario, N., & Odegaard, B. (2022). Color diversity judgments in peripheral vision: Evidence against "cost-free" representations. *PLOS One*, *17*(12), e0279686. https://doi.org/10.1371/journal.pone.0279686

[192] Plotinus. (Ennead, Volume V). Translated by A. H. Armstrong. Loeb Classical Library 444. Cambridge, MA: Harvard University Press, 1984.

[193] Henchsen, S. E. (1893). On the visual path and centre 1. *Brain*, *16*(1–2), 170–180. https://doi.org/10.1093/brain/16.1-2.170

[194] 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. (32893140). https://doi.org/10.1016/j.tics.2020.07.001

[195] Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. D. (2021). Deeply Felt Affect: The Emergence of Valence in Deep Active Inference. *Neural Computation*, *33*(2), 398–446. https://doi.org/10.1162/neco_a_01341

[196] 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

[197] Hobson, J. A., & Friston, K. J. (2014). Consciousness, dreams, and inference: The cartesian theatre revisited. *Journal of Consciousness Studies*, *21*(1–2), 6–32. Scopus.

[198] Hodgson, S. H. (1878). *The Philosophy of Reflection.* —. Longmans.

[199] Hodson, S. H. (1900). Perception of change and duration-a reply. *Mind*, *9*(34), 240–243.

[200] Hoerl, C. (2009). Time and tense in perceptual experience. *Philosophers' Imprint*, *9*, 1–18.

[201] Hoerl, C. (2013). ”A succession of feelings, in and of itself, is not a feeling of succession'. *Mind*, *122*(486), 373–417.

[202] 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

[203] Hogendoorn, H. (2020). Motion Extrapolation in Visual Processing: Lessons from 25 Years of Flash-Lag Debate. *The Journal of Neuroscience*, *40*(30), 5698–5705. https://doi.org/10.1523/JNEUROSCI.0275-20.2020

[204] 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

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

[206] Hohwy, J. (2013a). Action and expected experience. In J. Hohwy (Ed.), *The Predictive Mind* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.003.0005

[207] Hohwy, J. (2013b). Perceptual unity in action. In J. Hohwy (Ed.), *The Predictive Mind* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.003.0011

[208] Hohwy, J. (2013c). *The predictive mind*. OUP Oxford.

[209] Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701. https://doi.org/10.1016/j.cognition.2008.05.010

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

[211] Holcombe, A., & Corbett, J. (2023). *Temporal errors: Researchers should stop studying the flash-lag effect*. OSF. https://doi.org/10.31234/osf.io/swzr7

[212] Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, *13*(5), 216–221. https://doi.org/10.1016/j.tics.2009.02.005

[213] Holcombe, A. O. (2014). Are there cracks in the facade of continuous visual experience? In *Subjective time: The philosophy, psychology, and neuroscience of temporality* (pp. 179–198). Boston Review. https://doi.org/10.7551/mitpress/8516.001.0001

[214] 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

[215] Holcombe, A. O., & Corbett, J. J. (2023). *Temporal errors: Researchers should stop studying the flash-lag effect* (Swzr7_v1). PsyArXiv. https://doi.org/10.31234/osf.io/swzr7

[216] Holcombe, A. O., Treisman, A., & Kanwisher, N. (2001). The midstream order deficit. *Perception & Psychophysics*, *63*(2), 322–329. https://doi.org/10.3758/BF03194472

[217] Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences*, *9*(1), 1–23. https://doi.org/10.1017/S0140525X00021269

[218] Hollingworth, A. (2008). Memory for real-world scenes. In *The Visual World in Memory*. Routledge.

[219] Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *The Quarterly Journal of Experimental Psychology Section A*, *58*(8), 1415–1433. https://doi.org/10.1080/02724980443000647

[220] Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception & Psychophysics*, *81*(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

[221] Hommel, B., & Colzato, L. S. (2015). Learning from history: The need for a synthetic approach to human cognition. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01435

[222] Hommel, B., & Wiers, R. W. (2017). Towards a Unitary Approach to Human Action Control. *Trends in Cognitive Sciences*, *21*(12), 940–949. https://doi.org/10.1016/j.tics.2017.09.009

[223] Howe, P. D. L., & Holcombe, A. O. (2012). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of Vision*, *12*(13), 10. https://doi.org/10.1167/12.13.10

[224] Hsieh, L.-T., Ekstrom, A. D., & Ranganath, C. (2011). Neural oscillations associated with item and temporal order maintenance in working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(30), 10803–10810. https://doi.org/10.1523/JNEUROSCI.0828-11.2011

[225] Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

[226] Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, *12*(5), 822–851. https://doi.org/10.3758/bf03196775

[227] 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

[228] 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

[229] Hume, D. (1739). *A Treatise of Human Nature*. Oxford: Clarendon Press, 1896.

[230] Hurley, S. L. (1996). Unity and Objectivity. In C. Peacocke (Ed.), *Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind* (pp. 49–77). British Academy.

[231] Husserl, E. (1991). *On the Phenomenology of the Consciousness of Internal Time (1893-1917)*.

[232] Husserl, E. G. (1991). *On the Phenomenology of the Consciousness of Internal Time (1893-1917). Translated by John Barnett Brough*. Kluwer Academic Publishers.

[233] 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

[234] 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

[235] Iwasaki, S. (1993). Spatial attention and two modes of visual consciousness. *Cognition*, *49*(3), 211–233. https://doi.org/10.1016/0010-0277(93)90005-G

[236] 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

[237] Jackendoff, R. (1987). *Consciousness and the computational mind* (pp. xvi, 356). The MIT Press.

[238] Jackson, F. (1982). Epiphenomenal Qualia. Philosophical Quarterly, 32(127), 127–136. https://doi.org/10.2307/2960077

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

[240] Jacob, J., Breitmeyer, B. G., & Treviño, M. (2013). Tracking the first two seconds: Three stages of visual information processing? *Psychonomic Bulletin & Review*, *20*(6), 1114–1119. https://doi.org/10.3758/s13423-013-0482-4

[241] James, W. (1886). The perception of time. *Journal of Speculative Philosophy*, *20*(4), 374–407.

[242] James, W. (1890). *The Principles of Psychology* (Vol. 1, Issue n/a, p. 284). Dover Publications.

[243] Jertberg, R. M., Begeer, S., Geurts, H. M., Chakrabarti, B., & Van der Burg, E. (2024). Perception of temporal synchrony not a prerequisite for multisensory integration. *Scientific Reports*, *14*(1), 4982. https://doi.org/10.1038/s41598-024-55572-x

[244] 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

[245] Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology: CB*, *16*(5), 472–479. https://doi.org/10.1016/j.cub.2006.01.032

[246] 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

[247] Jones, L. A., & Wearden, J. H. (2003). More is not necessarily better: Examining the nature of the temporal reference memory component in timing. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, *56B*(4), 321–343. https://doi.org/10.1080/02724990244000287

[248] Kant, I. (1929). *Critique of Pure Reason (1781-1787), Trans. Kemp Smith*. MacMillan.

[249] Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*(4), 323–343. https://doi.org/10.1007/s00422-018-0753-2

[250] Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. Cognitive Psychology, 52(4), 346–368. https://doi.org/10.1016/j.cogpsych.2005.12.001

[251] Keller, A. (2016). The distinction between perception and judgment, if there is one, is not clear and intuitive. *Behavioral and Brain Sciences*, *39*, e249. https://doi.org/10.1017/S0140525X15002769

[252] Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (1999). Attention without awareness in blindsight. Proceedings. Biological Sciences, 266(1430), 1805–1811. https://doi.org/10.1098/rspb.1999.0850

[253] Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. Neuropsychologia, 42(6), 831–835. https://doi.org/10.1016/j.neuropsychologia.2003.11.001

[254] 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

[255] Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2009). Perception and hierarchical dynamics. *Frontiers in Neuroinformatics*, *3*, 20. https://doi.org/10.3389/neuro.11.020.2009

[256] 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

[257] 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

[258] Kiss, M., & Eimer, M. (2008). ERPs reveal subliminal processing of fearful faces. *Psychophysiology*, *45*(2), 318–326. https://doi.org/10.1111/j.1469-8986.2007.00634.x

[259] Klein, C. (2018). What Do Predictive Coders Want? *Synthese*, *195*(6), 2541–2557. https://doi.org/10.1007/s11229-016-1250-6

[260] 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

[261] Knotts, J. D., Odegaard, B., Lau, H., & Rosenthal, D. (2019). Subjective inflation: Phenomenology's get-rich-quick scheme. *Current Opinion in Psychology*, *29*, 49–55. https://doi.org/10.1016/j.copsyc.2018.11.006

[262] Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. https://doi.org/10.1016/j.tics.2006.10.012

[263] 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

[264] Köhler, W. (1947). *Gestalt psychology; an introduction to new concepts in modern psychology, Rev. Ed*. Liveright.

[265] Kohlrausch, A., van Eijk, R., Juola, J. F., Brandt, I., & van de Par, S. (2013). Apparent causality affects perceived simultaneity. *Attention, Perception & Psychophysics*, *75*(7), 1366–1373. https://doi.org/10.3758/s13414-013-0531-0

[266] 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

[267] Kolers, P. A., & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Res*, *16*(4), 329–335. (941407). https://doi.org/10.1016/0042-6989(76)90192-9

[268] Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception*, *33*(9), 1049–1060. https://doi.org/10.1068/p5169

[269] Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. Trends in Cognitive Sciences, 14(7), 301–307. https://doi.org/10.1016/j.tics.2010.04.006

[270] Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, *18*(9), 457–464. https://doi.org/10.1016/j.tics.2014.05.008

[271] 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

[272] Lambert, A. J., & Wootton, A. (2017). The time-course of activation in the dorsal and ventral visual streams during landmark cueing and perceptual discrimination tasks. *Neuropsychologia*, *103*, 1–11. https://doi.org/10.1016/j.neuropsychologia.2017.07.002

[273] Lamme, V. A. F. (2001). Blindsight: The role of feedforward and feedback corticocortical connections. *Acta Psychologica*, *107*(1–3), 209–228. https://doi.org/10.1016/s0001-6918(01)00020-8

[274] Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn Sci*, *10*(11), 494–501. (16997611). https://doi.org/10.1016/j.tics.2006.09.001

[275] Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. Scopus. https://doi.org/10.1016/S1364-6613(02)00013-X

[276] Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–220. https://doi.org/10.1080/17588921003731586

[277] 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. Scopus. https://doi.org/10.1016/S0166-2236(00)01657-X

[278] 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

[279] Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

[280] Le Poidevin, R. (2007). *The Images of Time: An Essay on Temporal Representation*. Oxford University Press.

[281] 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

[282] Lee, G. (2014). Temporal Experience and the Temporal Structure of Experience. *Philosophers' Imprint*, *14*.

[283] Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/JOSAA.20.001434

[284] Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, *4*(4), 501–506. https://doi.org/10.3758/BF03214339

[285] Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

[286] 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

[287] Li, Y., Zhong, W., Chen, Q., & Mo, L. (2013). Object detection is completed earlier than object categorization: Evidence from LRP and N200. *Psychophysiology*, *50*(12), 1255–1262. https://doi.org/10.1111/psyp.12131

[288] Libet, B. (1989). The timing of a subjective experience. *Behavioral and Brain Sciences*, *12*(1), 183–185. Cambridge Core. https://doi.org/10.1017/S0140525X00024912

[289] Libet, B. W. (1978). Neuronal vs. Subjective timing for a conscious sensory experience. In P. A. Buser & A. Rougeul-Buser (Eds.), *Cerebral Correlates of Conscious Experience*. Elsevier.

[290] 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

[291] Linares, D., & Holcombe, A. O. (2014). Differences in perceptual latency estimated from judgments of temporal order, simultaneity and duration are inconsistent. *I-Perception*, *5*(6), 559–571. https://doi.org/10.1068/i0675

[292] Litwin, P., & Miłkowski, M. (2020). Unification by Fiat: Arrested Development of Predictive Processing. *Cognitive Science*, *44*(7), e12867. https://doi.org/10.1111/cogs.12867

[293] Locke, J. (1689). *An essay concerning human understanding* (Vol. 3, Issue 5, pp. 601–605). Oxford University Press.

[294] Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A Psychophysical Investigation of Differences between Synchrony and Temporal Order Judgments. *PLOS one*, *8*(1), e54798. https://doi.org/10.1371/journal.pone.0054798

[295] Lupyan, G. (2015). Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems. *Review of Philosophy and Psychology*, *6*(4), 547–569. https://doi.org/10.1007/s13164-015-0253-4

[296] Mack, A., & Rock, I. (1998). *Inattentional blindness* (pp. xiv, 273). The MIT Press.

[297] Marcel, A. J. (1983). Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition. Cognitive Psychology, 15, 197–237.

[298] Marchetti, G. (2012). Against the View that Consciousness and Attention are Fully Dissociable. *Frontiers in Psychology*, *3*, 36. https://doi.org/10.3389/fpsyg.2012.00036

[299] Marois, R., Yi, D.-J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron*, *41*(3), 465–472. https://doi.org/10.1016/s0896-6273(04)00012-1

[300] Marr, D. (1982). *Vision* (Issue 3). W. H. Freeman.

[301] 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

[302] Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

[303] Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology*, *68*(4), 1332–1344. https://doi.org/10.1152/jn.1992.68.4.1332

[304] Mazzi, C., Bagattini, C., & Savazzi, S. (2016). Blind-Sight vs. Degraded-Sight: Different Measures Tell a Different Story. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.00901

[305] 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

[306] McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *The Journal of the Acoustical Society of America*, *77*(2), 678–685. https://doi.org/10.1121/1.392336

[307] McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

[308] McWalter, R., & McDermott, J. H. (2019). Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nature Communications*, *10*(1), 5096. https://doi.org/10.1038/s41467-019-12893-0

[309] Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLOS one*, *8*(8), e71608. https://doi.org/10.1371/journal.pone.0071608

[310] Mellor, D. H. (1981). *Real Time*. Cambridge University Press.

[311] Merikle, P. M., & Daneman, M. (1998). Psychological investigations of unconscious perception. *Journal of Consciousness Studies*, *5*(1), 5–18.

[312] Michel, M., & Lau, H. (2021). Is blindsight possible under signal detection theory? Comment on Phillips (2021). Psychological Review, 128(3), 585–591. https://doi.org/10.1037/rev0000266

[313] Miłkowski, M., & Litwin, P. (2022). Testable or bust: Theoretical lessons for predictive processing. *Synthese*, *200*(6), 462. https://doi.org/10.1007/s11229-022-03891-9

[314] Miller, I. (1984). *Husserl, Perception, And Temporal Awareness*. MIT Press.

[315] Miller, K., Holcombe, A., & Latham, A. J. (2020). Temporal phenomenology: Phenomenological illusion versus cognitive error. *Synthese*, *197*(2), 751–771.

[316] 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.

[317] Mole, C. (2011). *Attention is Cognitive Unison: An Essay in Philosophical Psychology*. Oxford University Press.

[318] Monno, I., Aufschnaiter, S., Ehret, S., Kiesel, A., Poljac, E., & Thomaschke, R. (2022). Time-based task expectancy: Perceptual task indicator expectancy or expectancy of post-perceptual task components? *Psychological Research*, *86*(5), 1665–1682. https://doi.org/10.1007/s00426-021-01588-1

[319] Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience*, *8*(7), 950–954. https://doi.org/10.1038/nn1488

[320] 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

[321] 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.

[322] Muth, F. V., Wirth, R., & Kunde, W. (2022). Temporal Binding in Multi-Step Action-Event Sequences is Driven by Altered Effect Perception. *Consciousness and Cognition*, *99*, 103299. https://doi.org/10.1016/j.concog.2022.103299

[323] Naccache, L. (2018). Why and how access-consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755), 20170357. https://doi.org/10.1098/rstb.2017.0357

[324] Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science*, *13*(5), 416–424. https://doi.org/10.1111/1467-9280.00474

[325] Neumann, O., & Klotz, W. (1994). Motor responses to nonreportable, masked stimuli: Where is the limit of direct parameter specification? In Attention and performance 15: Conscious and nonconscious information processing (pp. 123–150). The MIT Press.

[326] Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, *370*(6487), 256–257. https://doi.org/10.1038/370256b0

[327] 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

[328] Ninokura, Y., Mushiake, H., & Tanji, J. (2004). Integration of Temporal Order and Object Information in the Monkey Lateral Prefrontal Cortex. *Journal of Neurophysiology*, *91*(1), 555–560. https://doi.org/10.1152/jn.00694.2003

[329] 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

[330] 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. Cambridge Core. https://doi.org/10.1017/CBO9780511750540.017

[331] Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, *7*(1–3), 93–106. https://doi.org/10.1080/135062800394702

[332] Odegaard, B., Knight, R. T., & Lau, H. (2017). *Should a few null findings falsify prefrontal theories of conscious perception?* (p. 122267). bioRxiv. https://doi.org/10.1101/122267

[333] Öğmen, H., Ekiz, O., Huynh, D., Bedell, H. E., & Tripathy, S. P. (2013). Bottlenecks of Motion Processing during a Visual Glance: The Leaky Flask Model. *PLOS one*, *8*(12), e83671. https://doi.org/10.1371/journal.pone.0083671

[334] Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*, 145–175. https://doi.org/10.1023/A:1011139631724

[335] Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2

[336] O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, *24*(5), 939–973; discussion 973-1031. https://doi.org/10.1017/s0140525x01000115

[337] Otten, M., Pinto, Y., Paffen, C. L. E., Seth, A. K., & Kanai, R. (2017). The Uniformity Illusion. *Psychological Science*, *28*(1), 56–68. https://doi.org/10.1177/0956797616672270

[338] Overgaard, M., Fehl, K., Mouridsen, K., Bergholt, B., & Cleeremans, A. (2008). Seeing Without Seeing? Degraded Conscious Vision in a Blindsight Patient. *PLOS one*, *3*, e3028. https://doi.org/10.1371/journal.pone.0003028

[339] Parr, T., Corcoran, A. W., Friston, K. J., & Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, *2019*(1), niz012. https://doi.org/10.1093/nc/niz012

[340] Parr, T., Da Costa, L., & Friston, K. (2019). Markov blankets, information geometry and stochastic thermodynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *378*(2164), 20190159. https://doi.org/10.1098/rsta.2019.0159

[341] 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

[342] Patel, S. S., Ogmen, H., Bedell, H. E., & Sampath, V. (2000). Flash-lag effect: Differential latency, not postdiction. *Science*, *290*(5494), 1051. (11184992). https://doi.org/10.1126/science.290.5494.1051a

[343] Paton, J. J., & Buonomano, D. V. (2018). The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions. *Neuron*, *98*(4), 687–705. (29772201). https://doi.org/10.1016/j.neuron.2018.03.045

[344] Persaud, N., & Cowey, A. (2008). Blindsight is unlike normal conscious vision: Evidence from an exclusion task. Consciousness and Cognition: An International Journal, 17(3), 1050–1055. https://doi.org/10.1016/j.concog.2007.10.002

[345] Peters, M. A. K., Ro, T., & Lau, H. (2016). Who's afraid of response bias? *Neuroscience of Consciousness*, *2016*(1), niw001. https://doi.org/10.1093/nc/niw001

[346] Phillips, I. (2008). Perceiving temporal properties. *European Journal of Philosophy*, *18*(2), 176–202.

[347] Phillips, I. (2011). Indiscriminability and experience of change. *Philosophical Quarterly*, *61*(245), 808–827.

[348] Phillips, I. (2014a). Experience of and in Time. *Philosophy Compass*, *9*(2), 131–144.

[349] Phillips, I. (2014b). The Temporal Structure of Experience. In *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality, Edited by Valtteri Arstila, Dan Lloyd*. MIT Press. https://direct.mit.edu/books/edited-volume/3048/chapter/83274/The-Temporal-Structure-of-Experience

[350] Phillips, I. (2021). Blindsight is qualitatively degraded conscious vision. Psychological Review, 128(3), 558–584. https://doi.org/10.1037/rev0000254

[351] 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. (31279964). https://doi.org/10.1016/j.concog.2019.04.006

[352] Pitts, M. A., Lutsyshyna, L. A., & Hillyard, S. A. (2018). The relationship between attention and consciousness: An expanded taxonomy and implications for 'no-report' paradigms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170348. https://doi.org/10.1098/rstb.2017.0348

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

[354] Posner, M. I. (1994). Attention: The mechanisms of consciousness. Proceedings of the National Academy of Sciences of the United States of America, 91(16), 7398–7403. https://doi.org/10.1073/pnas.91.16.7398

[355] Pratt, H. (2011). Sensory ERP Components. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0050

[356] Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oup Usa.

[357] Prinz, J. J. (2010). When is Perception Conscious? In B. Nanay (Ed.), *Perceiving the World* (pp. 310–332). Oxford University Press.

[358] Prosser, S. (2016). *Experiencing Time*. Oxford University Press UK.

[359] Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197. https://doi.org/10.1163/156856888x00122

[360] Railo, H., Piccin, R., & Lukasik, K. M. (2021). Subliminal perception is continuous with conscious vision and can be predicted from prestimulus electroencephalographic activity. *The European Journal of Neuroscience*, *54*(3), 4985–4999. https://doi.org/10.1111/ejn.15354

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

[362] Rammsayer, T. H. (1999). Neuropharmacological Evidence for Different Timing Mechanisms in Humans. *The Quarterly Journal of Experimental Psychology Section B*, *52*(3), 273–286. https://doi.org/10.1080/713932708

[363] Ransom, M., Fazelpour, S., Markovic, J., Kryklywy, J., Thompson, E. T., & Todd, R. M. (2020). Affect-biased attention and predictive processing. *Cognition*, *203*, 104370. https://doi.org/10.1016/j.cognition.2020.104370

[364] 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

[365] Rashbrook, O. (2013). An Appearance of Succession Requires a Succession of Appearances. *Philosophy and Phenomenological Research*, *87*(3), 584–610. https://doi.org/10.1111/j.1933-1592.2012.00602.x

[366] Rashbrook, O. (2013). Diachronic and synchronic unity. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *164*(2), 465–484.

[367] Rashbrook-Cooper, O. (2016). *The Stream of Consciousness: A Philosophical Account* (pp. 117–134). https://doi.org/10.1007/978-3-319-22195-3_7

[368] Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*, *93*(2), 180–206.

[369] Reid, T. (1785). *Essays on the Intellectual Powers of Man* (Issue 66). Cambridge University Press.

[370] Rensink, R. A. (2000). The Dynamic Representation of Scenes. *Visual Cognition*, *7*(1/2/3), 17–42.

[371] Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

[372] 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

[373] Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings. Biological Sciences*, *278*(1710), 1314–1322. https://doi.org/10.1098/rspb.2010.1737

[374] Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences*, *114*(2), 412–417. https://doi.org/10.1073/pnas.1610706114

[375] Roberts, B. M., Libby, L. A., Inhoff, M. C., & Ranganath, C. (2018). Brain activity related to working memory for temporal order and object information. *Behavioural Brain Research*, *354*, 55–63. https://doi.org/10.1016/j.bbr.2017.05.068

[376] Robertson, L. C. (2003). BINDING, SPATIAL ATTENTION AND PERCEPTUAL AWARENESS. *Nature Reviews. Neuroscience*, *4*(2), 93. https://doi.org/10.1038/nrn1030

[377] Robinson, W. S. (2004). *Understanding Phenomenal Consciousness*. Cambridge University Press.

[378] Rohde, M., Greiner, L., & Ernst, M. O. (2014). Asymmetries in visuomotor recalibration of time perception: Does causal binding distort the window of integration? *Acta Psychologica*, *147*, 127–135. https://doi.org/10.1016/j.actpsy.2013.07.011

[379] Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, *12*(6), 852–877. https://doi.org/10.1080/13506280444000553

[380] Ruess, M., Thomaschke, R., Haering, C., Wenke, D., & Kiesel, A. (2018). Intentional binding of two effects. *Psychological Research*, *82*(6), 1102–1112. https://doi.org/10.1007/s00426-017-0892-4

[381] Ruess, M., Thomaschke, R., & Kiesel, A. (2017). The time course of intentional binding. *Attention, Perception & Psychophysics*, *79*(4), 1123–1131. https://doi.org/10.3758/s13414-017-1292-y

[382] 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

[383] 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

[384] Safavi, S., Kapoor, V., Logothetis, N. K., & Panagiotaropoulos, T. I. (2014). Is the frontal lobe involved in conscious perception? *Frontiers in Psychology*, *5*(SEP). Scopus. https://doi.org/10.3389/fpsyg.2014.01063

[385] Sahraie, A., Hibbard, P. B., Trevethan, C. T., Ritchie, K. L., & Weiskrantz, L. (2010). Consciousness of the first order in blindsight. *Proceedings of the National Academy of Sciences*, *107*(49), 21217–21222. https://doi.org/10.1073/pnas.1015652107

[386] Sampanes, A. C., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research*, *48*(21), 2275–2283. https://doi.org/10.1016/j.visres.2008.07.011

[387] Sanabria, F. (2020). Internal-Clock Models and Misguided Views of Mechanistic Explanations: A Reply to Eckard & Lattal (2020). *Perspectives on Behavior Science*, *43*(4), 779–790. https://doi.org/10.1007/s40614-020-00268-6

[388] Schenk, T., & Zihl, J. (1997). Visual motion perception after brain damage: I. Deficits in global motion perception. *Neuropsychologia*, *35*(9), 1289–1297. https://doi.org/10.1016/s0028-3932(97)00004-3

[389] Schlag, J., & Schlag-Rey, M. (2002). Through the eye, slowly: Delays and localization errors in the visual system. *Nature Reviews. Neuroscience*, *3*(3), 191–215. https://doi.org/10.1038/nrn750

[390] Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. (2008). Feedforward and recurrent processing in scene segmentation: Electroencephalography and functional magnetic resonance imaging. *J Cogn Neurosci*, *20*(11), 2097–2109. (18416684). https://doi.org/10.1162/jocn.2008.20142

[391] 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

[392] Schwarz, K. A., Weller, L., Pfister, R., & Kunde, W. (2019). Connecting action control and agency: Does action-effect binding affect temporal binding? *Consciousness and Cognition*, *76*, 102833. https://doi.org/10.1016/j.concog.2019.102833

[393] Schwarz, M. A., Winkler, I., & Sedlmeier, P. (2013). The heart beat does not make us tick: The impacts of heart rate and arousal on time perception. *Attention, Perception & Psychophysics*, *75*(1), 182–193. https://doi.org/10.3758/s13414-012-0387-8

[394] Schwitzgebel, E. (2009). When Our Eyes Are Closed, What, If Anything, Do We Visually Experience? In E. Schwitzgebel (Ed.), *Draft Available on Author's Homepage; Final Version in 2011 Monograph*.

[395] Senkowski, D., Saint-Amour, D., Gruber, T., & Foxe, J. J. (2008). Look who's talking: The deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *NeuroImage*, *43*(2), 379–387. https://doi.org/10.1016/j.neuroimage.2008.06.046

[396] Sereno, M. I., McDonald, C. T., & Allman, J. M. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *4*(6), 601–620. https://doi.org/10.1093/cercor/4.6.601

[397] Sergent, C., Ruff, C. C., Barbot, A., Driver, J., & Rees, G. (2011). Top–Down Modulation of Human Early Visual Cortex after Stimulus Offset Supports Successful Postcued Report. *Journal of Cognitive Neuroscience*, *23*(8), 1921–1934. https://doi.org/10.1162/jocn.2010.21553

[398] 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

[399] 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

[400] Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*(7), 439–452. https://doi.org/10.1038/s41583-022-00587-4

[401] Shen, M., Zhou, J., Gao, T., Liang, J., & Shui, R. (2007). The perceived position of a moving object is not the result of position integration. *Vision Research*, *47*(24), 3088–3095. https://doi.org/10.1016/j.visres.2007.08.012

[402] Shimojo, S. (2014). Postdiction: Its implications on visual awareness, hindsight, and sense of agency. *Front Psychol*, *5*, 196. (24744739). https://doi.org/10.3389/fpsyg.2014.00196

[403] Shimojo, Shinsuke. (2014). Postdiction: Its implications on visual awareness, hindsight, and sense of agency. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00196

[404] Siewert, C. (1998). *The Significance of Consciousness*. Princeton University Press.

[405] Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155. https://doi.org/10.1016/s1364-6613(00)01455-8

[406] Simons, D. J., & Ambinder, M. S. (2005). Change Blindness: Theory and Consequences. *Current Directions in Psychological Science*, *14*(1), 44–48. https://doi.org/10.1111/j.0963-7214.2005.00332.x

[407] Sincich, L. C., Park, K. F., Wohlgemuth, M. J., & Horton, J. C. (2004). Bypassing V1: A direct geniculate input to area MT. *Nature Neuroscience*, *7*(10), 1123–1128. https://doi.org/10.1038/nn1318

[408] Sinico, M. (1999). Benussi and the history of temporal displacement. *Axiomathes*, *10*(1), 75–93. https://doi.org/10.1007/BF02681817

[409] Slagter, H. A., Johnstone, T., Beets, I. A. M., & Davidson, R. J. (2010). Neural Competition for Conscious Representation across Time: An fMRI Study. *PLOS one*, *5*(5), e10556. https://doi.org/10.1371/journal.pone.0010556

[410] Soudry, Y., Lemogne, C., Malinvaud, D., Consoli, S.-M., & Bonfils, P. (2011). Olfactory system and emotion: Common substrates. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *128*(1), 18–23. https://doi.org/10.1016/j.anorl.2010.09.007

[411] 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

[412] Sprevak, M., & Smith, R. (2023). An Introduction to Predictive Processing Models of Perception and Decision-Making. *Topics in Cognitive Science*, tops.12704. https://doi.org/10.1111/tops.12704

[413] Stazicker, J. (2011). *Attention, Visual Consciousness and Indeterminacy*.

[414] Stein, T., Zwickel, J., Ritter, J., Kitzmantel, M., & Schneider, W. X. (2009). The effect of fearful faces on the attentional blink is task dependent. *Psychonomic Bulletin & Review*, *16*(1), 104–109. https://doi.org/10.3758/PBR.16.1.104

[415] 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

[416] Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *227*(2), 249–261. https://doi.org/10.1007/s00221-013-3507-3

[417] Stevenson, R. J. (2010). An initial evaluation of the functions of human olfaction. *Chemical Senses*, *35*(1), 3–20. https://doi.org/10.1093/chemse/bjp083

[418] Stewart, D. (1854). *The collected works of Dugald Stewart /* (W. Hamilton & J. Veitch, Eds.). T. Constable,.

[419] 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

[420] Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proc Biol Sci*, *268*(1462), 31–38. (12123295). https://doi.org/10.1098/rspb.2000.1326

[421] Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13. https://doi.org/10.1167/11.5.13

[422] Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). The Illusion of Uniformity Does Not Depend on the Primary Visual Cortex: Evidence From Sensory Adaptation. *I-Perception*, *9*(5), 2041669518800507. https://doi.org/10.1177/2041669518800728

[423] 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

[424] Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

[425] Sun, L., Frank, S. M., Hartstein, K. C., Hassan, W., & Tse, P. U. (2017). Back from the future: Volitional postdiction of perceived apparent motion direction. *Vision Research*, *140*, 133–139. https://doi.org/10.1016/j.visres.2017.09.001

[426] Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

[427] Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. Scopus. https://doi.org/10.1016/j.tics.2017.12.005

[428] Tanaka, T., Matsumoto, T., Hayashi, S., Takagi, S., & Kawabata, H. (2019). *What Makes Action and Outcome Temporally Close to Each Other: A Systematic Review and Meta-Analysis of Temporal Binding*. https://doi.org/10.1163/22134468-20191150

[429] Teichmann, L., Edwards, G., & Baker, C. I. (2021). Resolving visual motion through perceptual gaps. *Trends in Cognitive Sciences*, *25*(11), 978–991. https://doi.org/10.1016/j.tics.2021.07.017

[430] 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

[431] Thomas, N. J. T. (1999). Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content. *Cognitive Science*, *23*(2), 207–245. https://doi.org/10.1207/s15516709cog2302_3

[432] Thomaschke, R., Kunchulia, M., & Dreisbach, G. (2015). Time-based event expectations employ relative, not absolute, representations of time. *Psychonomic Bulletin & Review*, *22*(3), 890–895. https://doi.org/10.3758/s13423-014-0710-6

[433] 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

[434] Thorpe, S., Fize, D., & Marlot, C. (2007). Speed of Processing in the Human Visual System. *Nature*, *381*, 520–522. https://doi.org/10.1038/381520a0

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

[436] 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

[437] 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

[438] Trapp, S., Parr, T., Friston, K., & Schröger, E. (2021). The Predictive Brain Must Have a Limitation in Short-Term Memory Capacity. *Current Directions in Psychological Science*, *30*(5), 384–390. https://doi.org/10.1177/09637214211029977

[439] Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178. https://doi.org/10.1016/S0959-4388(96)80070-5

[440] Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron*, *24*(1), 105–110, 111–125. https://doi.org/10.1016/s0896-6273(00)80826-0

[441] Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

[442] Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

[443] Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock.” *Psychological Monographs: General and Applied*, *77*(13), 1–31. https://doi.org/10.1037/h0093864

[444] Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics*, *66*(7), 1171–1189. https://doi.org/10.3758/bf03196844

[445] Tsuchiya, N., & Koch, C. (2016). Chapter 5—The Relationship Between Consciousness and Top-Down Attention. In S. Laureys, O. Gosseries, & G. Tononi (Eds.), *The Neurology of Conciousness (Second Edition)* (pp. 71–91). Academic Press. https://doi.org/10.1016/B978-0-12-800948-2.00005-4

[446] 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

[447] Turner, W., Sexton, C., Johnson, P. A., Wilson, E. M., & Hogendoorn, H. (2025). Predictable motion is progressively extrapolated across temporally distinct processing stages in the human visual cortex. *PLOS Biology*, *23*(5), e3003189. https://doi.org/10.1371/journal.pbio.3003189

[448] Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging*, *31*(4), 380–389. https://doi.org/10.1037/pag0000094

[449] van Boxtel, J. J. A., Tsuchiya, N., & Koch, C. (2010). Opposing effects of attention and consciousness on afterimages. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8883–8888. https://doi.org/10.1073/pnas.0913292107

[450] Van de Cruys, S. (2017). *Affective Value in the Predictive Mind*. 0–0.

[451] Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

[452] Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, *55*(3), 1208–1218. https://doi.org/10.1016/j.neuroimage.2010.12.068

[453] van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics*, *70*(6), 955–968. https://doi.org/10.3758/PP.70.6.955

[454] van Rijn, H., Gu, B.-M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. *Advances in Experimental Medicine and Biology*, *829*, 75–99. https://doi.org/10.1007/978-1-4939-1782-2_5

[455] Van Volkinburg, H., & Balsam, P. (2014). Effects of Emotional Valence and Arousal on Time Perception. *Timing & Time Perception, 2*(3), 360–378. https://doi.org/10.1163/22134468-00002034

[456] van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

[457] Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research*, *185*(3), 521–529. https://doi.org/10.1007/s00221-007-1168-9

[458] Velmans, M. (1991). Is human information processing conscious? *Behavioral and Brain Sciences*, *14*(4), 651–669. https://doi.org/10.1017/S0140525X00071776

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

[460] Viera, G. A. (2019). The fragmentary model of temporal experience and the mirroring constraint. *Philosophical Studies*, *176*(1), 21–44. https://doi.org/10.1007/s11098-017-1004-4

[461] Volberg, G., & Thomaschke, R. (2017). Time-based expectations entail preparatory motor activity. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *92*, 261–270. https://doi.org/10.1016/j.cortex.2017.04.019

[462] Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659. https://doi.org/10.3758/bf03194427

[463] 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

[464] Vroomen, J., & de Gelder, B. (2004). Temporal Ventriloquism: Sound Modulates the Flash-Lag Effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 513–518. https://doi.org/10.1037/0096-1523.30.3.513

[465] Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

[466] Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

[467] Vul, E., & Rich, A. N. (2010). Independent sampling of features enables conscious perception of bound objects. *Psychological Science*, *21*(8), 1168–1175. https://doi.org/10.1177/0956797610377341

[468] Wallis, T. S., Funke, C. M., Ecker, A. S., Gatys, L. A., Wichmann, F. A., & Bethge, M. (2019). Image content is more important than Bouma's Law for scene metamers. *eLife*, *8*, e42512. https://doi.org/10.7554/eLife.42512

[469] Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research, Vision Research 50th Anniversary Issue: Part 1*, *51*(7), 718–737. https://doi.org/10.1016/j.visres.2010.08.004

[470] Wang, X., & Wöllner, C. (2020). Time as the Ink That Music Is Written With: A Review of Internal Clock Models and Their Explanatory Power in Audiovisual Perception. *Yearbook of Music Psychology*, 1–22. https://doi.org/10.5964/jbdgm.2019v29.67

[471] Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036-i

[472] 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

[473] White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review*, *116*(3), 580–601. https://doi.org/10.1037/a0016337

[474] White, P. A. (2012). The experience of force: The role of haptic experience of forces in visual perception of object motion and interactions, mental simulation, and motion-related judgments. *Psychological Bulletin*, *138*(4), 589–615. https://doi.org/10.1037/a0025587

[475] 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

[476] White, P. A. (2020). The perceived present: What is it, and what is it there for? *Psychonomic Bulletin & Review*, *27*(4), 583–601. https://doi.org/10.3758/s13423-020-01726-7

[477] White, P. A. (2023). Time marking in perception. Neuroscience & Biobehavioral Reviews, 146, 105043. https://doi.org/10.1016/j.neubiorev.2023.105043

[478] Wijntjes, M. W. A., & Rosenholtz, R. (2018). Context mitigates crowding: Peripheral object recognition in real-world images. *Cognition*, *180*, 158–164. https://doi.org/10.1016/j.cognition.2018.06.015

[479] 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

[480] Yamamoto, S., & Kitazawa, S. (2001). Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience*, *4*(7), Article 7. https://doi.org/10.1038/89559

[481] Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*(3), 295–340. https://doi.org/10.1016/0010-0285(92)90010-y

[482] 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

[483] 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

[484] Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67*(3), 531–544. https://doi.org/10.3758/BF03193329

[485] Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences*, *7*(5), 214–218. https://doi.org/10.1016/s1364-6613(03)00081-0

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

[487] 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

[488] Zeki, S., & Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1377), 1911–1914. https://doi.org/10.1098/rstb.1998.0343

[489] Zeki, S., & Bartels, A. (1999). Toward a theory of visual consciousness. *Consciousness and Cognition*, *8*(2), 225–259. https://doi.org/10.1006/ccog.1999.0390

[490] Zeki, S., & Bartels, A. (2009). The asynchrony of consciousness. *Proceedings. Biological Sciences / The Royal Society*, *265*, 1583–1585. https://doi.org/10.1098/rspb.1998.0475

[491] Zeleznikow-Johnston, A., Aizawa, Y., Yamada, M., & Tsuchiya, N. (2023). Are Color Experiences the Same across the Visual Field? *Journal of Cognitive Neuroscience*, *35*(4), 509–542. https://doi.org/10.1162/jocn_a_01962

[492] 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

[493] 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

[494] Zhou, Z., Chen, Y., & Fu, S. (2022). The effects of expectation on attention are dependent on whether expectation is on the target or on the distractor. *Acta Psychologica Sinica*, *54*(3), 221–235. https://doi.org/10.3724/SP.J.1041.2022.00221

[495] Zimmermann, E., Morrone, M. C., & Burr, D. C. (2014). Buildup of spatial information over time and across eye-movements. *Behavioural Brain Research*, *275*, 281–287. https://doi.org/10.1016/j.bbr.2014.09.013

[496] Zivony, A., & Eimer, M. (2024). Attention and feature binding in the temporal domain. *Psychonomic Bulletin & Review*, *31*(6), 2599–2610. https://doi.org/10.3758/s13423-024-02493-5