Acta Psychologica 220 (2021) 103403

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The extended present: an informational context for perception

Peter A. White

School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3YG, Wales, UK

ARTICLE INFO

Keywords:
Subjective present
Specious present
Hierarchical structure processing
Temporal information processing
Working memory

ABSTRACT

Several previous authors have proposed a kind of specious or subjective present moment that covers a few seconds of recent information. This article proposes a new hypothesis about the subjective present, renamed the extended present, defined not in terms of time covered but as a thematically connected information structure held in working memory and in transiently accessible form in long-term memory. The three key features of the extended present are that information in it is thematically connected, both internally and to current attended perceptual input, it is organised in a hierarchical structure, and all information in it is marked with temporal information, specifically ordinal and duration information. Temporal boundaries to the information structure are determined by hierarchical structure processing and by limits on processing and storage capacity. Supporting evidence for the importance of hierarchical structure analysis is found in the domains of music perception, speech and language processing, perception and production of goal-directed action, and exact arithmetical calculation. Temporal information marking is also discussed and a possible mechanism for representing ordinal and duration information on the time scale of the extended present is proposed. It is hypothesised that the extended present functions primarily as an informational context for making sense of current perceptual input, and as an enabler for perception and generation of complex structures and operations in language, action, music, exact calculation, and other domains.

It has been claimed that there is a kind of subjective or experienced present that encompasses information on a time scale of a few seconds. An example often taken is that of listening to a piece of music. This example appears to have originated with St. Augustine (Rovelli, 2018) and has been discussed by several authors since then (Clay, 1882; Durgin & Sternberg, 2002; Lloyd, 2012; Wittmann, 2011). To make the example more concrete, take the opening of Beethoven's fifth symphony (shown in reduced score in Fig. 1). When the eighth and last chord in Fig. 1 is heard, the argument is that it is heard not just as an isolated chord but in the context of the seven chords that preceded it. Although those seven chords are, strictly speaking, in the past and no longer perceived, they seem somehow to be still present in the mind, as a context that influences perception of the eighth chord. This need not be true of all recent perceptual information. If, during the first seven chords, the listener rubbed their arm and had tactile sensations of the rubbing, even if information about the rubbing was still held in memory at the time of the eighth chord, it would not influence how that chord is perceived, in the way that the first seven chords do. It is separate, not informationally bound with the eighth chord. The example is anecdotal and is just an aid to introduce the topic of the paper, but it serves to illustrate two important features. One is that there appears to be some kind of integration of recent historical information, and what is currently being perceived is perceived as in relation to that. The other is that the recent information in question is in working memory (WM) but is not the whole of what is in WM. The tactile sensations of rubbing could be held in WM just as the first seven chords of the symphony could be, but they are not part of the integrated set of information.

The general form of the previous proposals has been that there is global temporal integration that generates a coherent, overall sense of what is going on, on a time scale often said to be between 2 and 6 s (Fraisse, 1984; James, 1890; Montemayor & Wittmann, 2014; Pöppel, 1997, 2009; Wittmann, 1999, 2009, 2011, 2013; Wittmann & Pöppel, 1999/2000). For example, Wittmann (2009) wrote, “a perceptual mechanism seems to exist that integrates separate successive events into a unit or perceptual *gestalt*... We do not just perceive individual events in isolation, but automatically integrate them into perceptual units with a duration of approximately 2-3 s” (p. 1959). Research evidence cited in support of the proposals concerned automatic temporal integration in multiple domains all operating on approximately the same time scale (Montemayor & Wittmann, 2014; Pöppel, 1997, 2009; Wittmann, 1999, 2009, 2011). A recent review found that the research evidence did not support the proposals, showing that phenomena in different domains

---

E-mail address: whitepa@cardiff.ac.uk.

https://doi.org/10.1016/j.actpsy.2021.103403
Received 7 June 2021; Received in revised form 4 August 2021; Accepted 19 August 2021
Available online 26 August 2021

had different mean integration times, indicative of local, independent processes with domain-specific features (White, 2017).

It can, however, be argued that time scales and temporal features of integration are not the right kind of evidence; not only that, but some of the evidence called on in support of the proposals does not seem to be relevant to them. For example, Pöppel (1997, 2009) and Wittmann et al. (2011) argued that reversals of perception in bistable ambiguous figures such as the Necker cube occurred on the time scale of the subjective present. Explanations for perceptual reversals have been proposed that treat them as a local phenomenon, not connected to other kinds of processing (Furstenau, 2014; Lee, 2014), so it is likely that they are not relevant to the subjective present. So what does constitute the subjective present, if there is one? What kind of processing creates it? It is proposed here that there is a particular kind of integration of information that subserves the function of providing an informational context in which current perception is set. To distinguish this from the subjective present it will be termed the "extended present" (EP). The EP can encompass information in any sensory modality and of any kind: what marks the EP out is the particular ways in which information is bound into a coherent, organised, active structure.

1. Main features of the extended present

The EP is an information structure in WM. It should not be envisaged as a static, persisting representation; as will be shown, information in the EP is subject to constant revision because of the operation of several mechanisms, including changes in the focus of perceptual attention, entry of new perceptual information, integration of that new information with what is already there, activation of information stored in long-term memory (LTM), decay of information, depending on operation of information maintenance mechanisms such as rehearsal, and active removal of outdated information. At any given moment, the EP comprises a current product of perceptual processing (e.g. a musical chord or a word) as the temporal lead in a thematically connected and temporally differentiated structure of information housed in temporary store on the supra-second time scale. The information structure in the EP at a given moment is defined by thematic connections linking its contents, by temporal event boundaries, by temporal information, including information about the temporal order of input items of information and their individual durations, and by information about hierarchical structure and its boundaries. The amount of information in the structure at any time is limited by processing and storage capacity limitations and by identifiers of structural boundaries. Thus, the temporal extent of information in the structure may often approximate to a few seconds, but the EP is not defined in terms of time. "Temporal extent" does not mean the amount of time for which information persists in the EP. It refers to the temporal information that marks the most recent and oldest information currently in the EP at a given moment of time. That is, it is determined by temporal marking mechanisms, not by actual duration.

For initial orientation, Fig. 2 presents an abstract snapshot of a hypothetical moment in the EP, emphasizing the connectedness of information in it as well as hierarchical structure and temporal information. Both of those kinds of information will be analysed in more detail in sections to follow. The domain of information represented in the snapshot is not identified, to emphasise the domain-general nature of the EP. The word "snapshot" is used to make clear that the time referents in the figure do not indicate how long the information persists there: they are information about temporal features and they exist at a given moment, contemporaneously.

The EP also holds other, domain-specific kinds of information such as semantic and prosodic information. In the Beethoven example, the information currently in the EP about the music prior to the currently heard chord may include timbre, pitch, volume, tempo changes, and instrument identification information. Information from any sensory modality may be incorporated in the information structure, including exteroception, proprioception, and interoception, as may information about mental events such as imagery and the products of reasoning processes. The present account focusses mainly on exteroceptive perceptual information because that is where most of the relevant research has been carried out.

The information structure of the EP is constructed by means of three main kinds of processes. One is perceptual processing with information passing from perceptual processing either through sensory memory (Atkinson & Shiffrin, 1968; Coltheart, 1980; Ögmen & Herzog, 2016; Sligte, Vandenbroucke, Scholte, & Lamme, 2010; Sperling, 1960) to WM or directly to WM. This sequence of information processing is common to all kinds of information in WM so little more will be said about it in this paper. The second is spreading activation. In spreading activation, active and attended information in perceptual processing and WM activates further information from LTM, with activation being strongest for stored information that has the strongest associative bonds with the active information. Spreading activation can activate any kind of information that is stored in LTM, including all kinds of semantic information. For the EP, the main kind of information of interest is hierarchical structure information, so spreading activation will be further explained in the section on hierarchical structure processing. The third kind of process is time marking. Time marking involves temporal information being generated, read off, and attached to items of information in WM. How this is done will be further explained in the section on temporal information in the EP. The combination of these three kinds of processes generates a coherent body of information that exists at one time, in which individual items of information are labelled and related to other items of information in terms of location in time and in a hierarchical structure.

That can be contrasted with the classic stimulus presentation in experiments on STM capacity, a sequence of random digits. In such a presentation, each digit is an isolated stimulus and even their temporal order is irrelevant unless the instruction is to recall them in order. Such information may be held in WM but is not in the EP, because the individual items are not related to each other by temporal and hierarchical

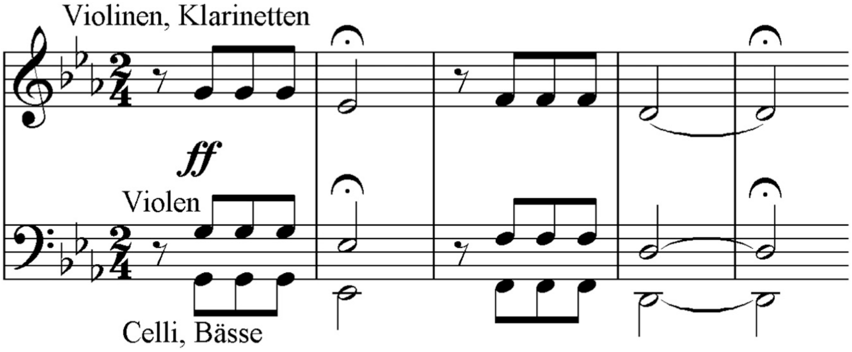

Fig. 1. First eight chords of Beethoven's fifth symphony (in reduced score).

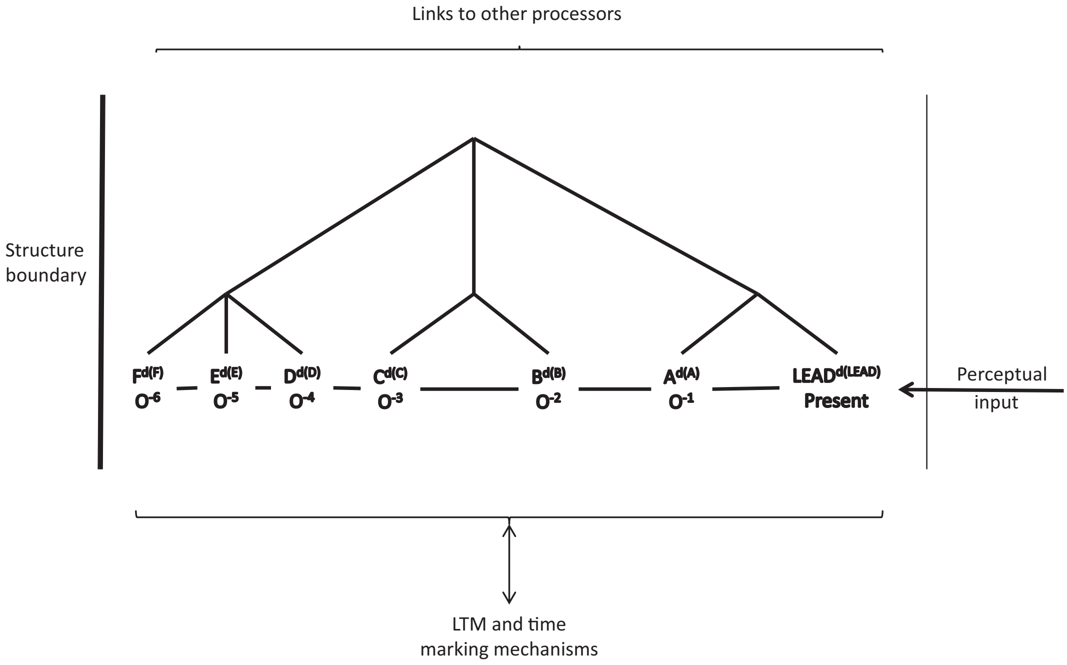

Fig. 2. Snapshot of a hypothetical moment in the EP. “LEAD” is the most recent perceptual object to enter the EP from perceptual processing. Letters A to F indicate successively more temporally distal objects. Superscripts “d(x)” are duration markers for individual objects. “Present” and “O^{-1}” to O^{-6}″ are ordinal time markers for individual objects. The tree structure above the objects is an abstract, simplified hierarchical structure. Vertical lines at left and right demarcate the temporal boundaries of the EP.

structure information. Any information in WM that is not connected by temporal (e.g. temporal order) information and hierarchical structure information is not in the EP. The EP does not necessarily go on all the time. It is there when the kinds of information that constitute it are in WM.

The main function of the EP is to provide an informational context that contributes to the interpretation of current attended perceptual input, thereby constituting an overall sense of what is going on. In addition, an important function of the EP is to hold integrated representations of information generated in hierarchical structure processing and temporal judgment mechanisms, in the service of further processing and action. Thus, information in the EP can be accessed and read off by other processing mechanisms, such as predictive mechanisms and action generation mechanisms.

To be clear, the purpose of this paper is not to say anything new about hierarchical structure processing in any of the domains considered here (music, language, action, and mathematics), nor is it to say anything new about time marking or timing mechanisms, or about event boundaries. It is to say something new about what the subjective present is. The proposal is completely different from any of the previous proposals about the subjective present. Surveys of research on hierarchical structure processing, timing mechanisms, and event boundaries are necessary to back up the proposal being made about what the subjective present really is. That is what the paper is about.

1.1. The EP and working memory

The counter-hypothesis to the present proposal would be that information is just stored temporarily in WM, and there is nothing distinctive about what is claimed as being the extended present. That possibility will now be evaluated. An early proposal of short-term memory as a unitary temporary limited capacity store (Atkinson & Shiffrin, 1968) was superseded by a multi-component model of WM, based on evidence for separate special-purpose stores (Baddeley, 2000; Baddeley & Hitch, 1974). The components in Baddeley's model of WM include a store of acoustic verbal information with a capacity of two or three items, now known as the phonological loop (Baddeley, 2000; Baddeley & Hitch, 1974), a store of visual information with a capacity of a single contemporaneous pattern, called the visuo-spatial sketchpad (Baddeley, 1983), a temporary holding area that integrates information from different sources, including information activated from LTM, called the episodic buffer (Baddeley, 2000), and a limited capacity attentional control system with flexible allocation of processing resources, the central executive (Baddeley & Hitch, 1974). Thus, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer are stores of information, and they are associated with a control system, the central executive, and with processes that operate on information in individual components of the model.

The EP may comprise information in any modality, so it would not be identified with either the phonological loop or the visuo-spatial sketchpad. It could be proposed, however, that it is just a specific kind of information that is housed in the episodic buffer. Baddeley (2000) characterised the episodic buffer as “a limited-capacity temporary storage system that is capable of integrating information from a variety of sources” (p. 421). The term “episodic” refers to the possibility that information in it is “integrated across space and potentially extended across time” (Baddeley, 2000, p. 421). It has a two-way link to episodic LTM and it may represent information in a range of codes.

In fact that proposal is not satisfactory. The EP is not the whole of what may be held in the episodic buffer. The EP is an information structure that is specifically associated with a set of generative processes, principally hierarchical structure analysis and time marking, that impose a specific kind of organization on the information that is held in it. There is no requirement in Baddeley's (2000) account of the episodic buffer that information in it must be organised by either hierarchical or temporal structure. Furthermore, information in the phonological loop and the visuo-spatial sketchpad may also be hierarchically organised, most obviously in the case of the phonological loop by syntactic structure, and information in the phonological loop must almost inevitably be marked with time order information. The EP is an information structure, not an information store; its contents may be distributed across more than one store in WM. It is defined in terms of the kind of information it contains and the processes that generate it, principally hierarchical structure analysis and time marking, not as a distinct store. The evidence for the phonological loop, the visuo-spatial sketchpad, and the episodic buffer as distinct stores does not imply that a defined kind of information structure could not be represented, partly or wholly, in any of them. Information can be related and integrated across stores.[1]

The point of the EP is to provide an informational context for what is currently being perceived. Thus, a chord in a piece of music that is being heard now is perceived and interpreted in terms of the retained information in the EP about preceding chords. An action being perceived now is perceived and interpreted in the informational context of preceding actions, for example as a step on the route to making a cup of coffee. That is not true of all information that is in WM. A digit that was presented in a digit span test five seconds ago and that is now in WM has no informational connection to what is being perceived now. A note in a piece of music that was heard five seconds ago does have an informational connection to what is being perceived now, and that informational connection plays a role in perceptual interpretation of the note that is being heard now. That is the functional significance of the EP, as it was supposed to be of the previous proposals of a subjective present, and it is what sets the EP apart, functionally speaking, from other information in WM.

The stores proposed in Baddeley's model are not the end of the story as far as WM is concerned. There is evidence for several other stores in WM each with their own capacity, including a store of semantic information, in particular integrated perceptual objects (Fougnie, Cormiea, & Alvarez, 2013; Joseph, Kumar, Husain, & Griffiths, 2015; Liesefeld & Müller, 2019; Loaiza & Camos, 2018; Martin & Romani, 1994; Mathias & von Kriegstein, 2014; Nishiyama, 2018; Shivde & Anderson, 2011), a tonal loop for musical information (Pechmann & Mohr, 1992; Schendel & Palmer, 2007; Schulze & Koelsch, 2012), and possibly separate stores for auditory features of musical stimuli such as pitch and timbre (Caclin & Tillmann, 2018). Information in multiple modalities may be stored in WM. Cowan (2017) made the point that there might be "a nonverbal auditory module, a haptic module, a semantic module, an orthographic module, and so on" (p. 1163). The "and so on" covers a multitude of possibilities: information from the many kinds of interoceptive sensors would be an example (Craig, 2009). Each store may have its own capacity, and decay of information in any store may be avoided by maintenance mechanisms such as rehearsal or by transfer in an accessible state to LTM (Norris, 2017). Information of any of these kinds could be held in the EP, further emphasizing that the EP is a kind of information structure, not a store.

WM operations often involve activation of information in LTM. Several studies have found evidence for a distinct store of semantic information in WM (Martin & Romani, 1994; Shivde & Anderson, 2011). The source of semantic information is LTM, so storage of semantic information indicates that LTM is involved in WM tasks. The classic example is research showing that sentences can be retained accurately despite having far longer duration than the 2 s limit on the phonological loop in the absence of rehearsal (Wingfield & Butterworth, 1984). Norris (2017) has argued that there may be no task involving WM in which LTM is not also involved: "sustained semantic activation is fed back to STM to aid retention. This is sometimes referred to as semantic binding" (p. 994). In some models, items initially entered into WM that are not the immediate attentional focus of an ongoing task may be assigned to LTM but in an activated, that is to say readily accessed, state (Cowan, 1995, 1999; Mallett & Lewis-Peacock, 2018; Oberauer, 2002). It will be shown that this is of direct relevance to the EP.

Processing functions associated with components of WM may play an important supportive role. For example, acoustic information decays in about two seconds (Baddeley & Hitch, 1974). That would severely limit the capacity of the EP and the complexity of the information structure of acoustic information that could be held in it. Acoustic information can be maintained in an active state by rehearsal utilising the phonological loop, however. Information can be temporarily fixed by consolidation, a function that prevents information from being disrupted by new perceptual input (Schurgin, 2018; Vogel et al., 2006). An additional mechanism for maintaining information in an active state is refreshing, which involves maintaining the focus of attention on the material in question or attentively reactivating it from LTM (Baddeley & Hitch, 2019; Loaiza & Camos, 2018; Nee & Jonides, 2013). These are examples of WM operations that can function to maintain information in the EP.

To summarise, the EP is not to be identified with or located in a single store in WM. What marks out the EP is not which store it is held in, but the kind of processing that generates it; that is, it is marked out as an information structure of a particular kind, principally involving hierarchical organization and time marking, and not as a location where information may be found. It can be distributed across various stores in WM depending on what sort of information is in it at any given time. Information temporarily activated from LTM forms an important part of the EP, because such information adds semantic content to the representation and it allows for more complex information structures to be put together.

The remainder of the paper has the following plan. The next section elucidates how hierarchical structure information in the EP is constructed, seeks evidence for hierarchical structure processing in the four domains of music, language, action, and arithmetic, and assesses the case for domain-general hierarchical structure processing. The section after that elucidates how temporal information is generated and attached to items of information in the EP. The section after that considers how boundaries on information structures in the EP are generated. Later sections consider the possible functional significance of the EP, compare the EP with "subjective present" proposals, and assess whether there could be a basic temporal unit of information in the EP.

2. Hierarchical structure information

When the second chord of Beethoven's fifth symphony is presented, the EP contains information about its thematic connection to the first chord and relevant temporal information, mainly the duration of each chord and their order in time. Thematic connection means that the items are not represented in the EP as isolated individual occurrences but as forming a kind of whole, of which each item is an integrated part. At that stage, there is no hierarchical structure in the information. But, as the music continues, a hierarchical structure begins to emerge. The main proposed function of the EP is to provide an informational context that contributes to the interpretation of current perceptual input, and an important part of that is locating current perceptual input in a temporal hierarchical structure.

To generate a hierarchical structure, two or more elements must be combined to make a higher-order element and then related to at least one other element. To take an example from language, "a determiner (D) *the* and a noun (N) *man* are combined to create a higher-order element, a determiner phrase (DP), *the man*. This DP can combine with a verb (V) *run* to make up a next higher-order element, a sentence (S), *the man runs*" (Friederici, 2019, p. 2). That is an example of Merge, argued by Chomsky (1995) to be the basic computation to generate a syntactic hierarchy. Increasingly complex hierarchical structures can be generated by successive iterations of the same computational operation. An equivalent principle of combination can operate in any domain of

---

[1] How does the EP relate to the central executive in Baddeley's model? First, the EP and the central executive are different and in fact doubly dissociated: the EP is an information structure held in WM whereas the central executive has no storage capacity of its own (Baddeley, 2000; Baddeley & Hitch, 1974); and the EP has no executive function whereas executive function is the defining characteristic of the central executive. If there is such a thing as the central executive, then the information that gets into the EP may be determined in part by the operation of the central executive, and the central executive may also operate on information held in the EP.

cognition: the kinds of units that are involved may differ but the basic computational procedure of unitisation and integration is common to all domains. In the example of Beethoven's fifth symphony, the first three chords in Fig. 1, presented on their own, may be represented as thematically connected and unitised and with temporal information such as ordinal relation but no more. When combined with the fourth note, they become an element in a higher-order musical structure. The particular structure is not established until the fourth chord is heard. Just as reversing the order of the units in *the man runs* would create a different structure, *runs the man*, with a different meaning (in English, *man* is now in the usual position of the grammatical object), so reversing the order of the fourth chord and the group of the first three would result in a different musical structure.

Hierarchical structures have other common features. Embedding refers to relationships of dominance and subordination between units or sets of units. As an example, "when the word 'film' is embedded in 'committee' to form [[film] committee], it refers to a kind of committee, not a kind of film" (Martins et al., 2020, p. 2). Recursion involves embedding an expression of some kind within an expression of the same kind. In mathematics recursion involves describing numbers or other mathematical expressions in terms of the numbers or expressions that come before them in a series. For instance, "the natural numbers are described by the recursive function $N_{\mathrm{i}} = N_{i\text{-}1} + 1$, which generates the infinite set {1, 2, 3,...}" (Martins et al., 2020, p. 2). In language, recursion involves putting a structure such as a phrase within a structure of the same kind. Embedding and recursion can be used to generate hierarchical information structures of indefinite size.

Temporality is a feature of all hierarchical structures in the EP but not all hierarchical structures are temporal hierarchies. To borrow an example from Fitch and Martins (2014), there can be part-whole (meronomic) hierarchies, such as is the case for the human face. The face as a whole contains identifiable parts (e.g. eyes, nose) and a part may have parts as well, such as iris and pupil in the eye, so it could be said that the face has a meronomic hierarchical structure with at least three levels. There are also social and institutional hierarchies such as power hierarchies (Martins, Muršic, Oh, & Fitch, 2015). The present concern is specifically with structural hierarchies in which temporal order makes a difference to the structure, with consequences for semantics. Thus, a temporal hierarchical structure can be analysed as a tree structure in which each element is described in terms of its temporal relationship with the other elements as well as in terms of hierarchical relationships.

2.1. *Constructing hierarchical structure information in the extended present*

As Tillmann (2012) pointed out, "implicit learning enables perceivers to learn regularities in their environment through mere exposure to materials that obey the rules of a given system" (p. 569). Thus, structural regularities in actions, language, and music may be implicitly learned. The regular temporal associations between structural features result in representations of structural information in LTM with strong associative bonds linking their components. In the example of music, it has been shown in many studies (reviewed by Tillmann, 2012) that structural knowledge activated by musical stimuli functions as a set of expectations for forthcoming events, and events that confirm expectations are then processed more quickly than unexpected events. Learning of structural information has been demonstrated in infants aged about six months (Daum, Prinz, & Aschersleben, 2009; Falck-Ytter, Gredebäck, & von Hofsten, 2006; Luo, 2010; Maffongelli, Antognini, & Daum, 2018). Thus, even early in life, there is a resource of stored information about hierarchical structure that can be activated and put to work in the interpretation of perceived behaviour and also in the construction of behaviour in various domains. This section focusses on domain-general aspects of the construction of hierarchical structure information in the EP, based on information activated from LTM.

Stored information is involved in processing from early in perception. Take the example of the word "king" presented as a stimulus. This is a collection of acoustic features (if an auditory stimulus) or of lines, angles, and spatial relations between them (if a visual stimulus). In vision, early stages of perceptual processing register individual features which are then assembled to form a perceptual object, which means that the object is individuated (separated from other sensory input and treated as a bounded unit) and identified. Identification involves activated semantic information. Putting it rather simply, there is a feedforward sweep of information through early stages of perceptual processing that analyses the surface features of the stimulus. That then triggers feedback (or recurrent or reentrant) processing in which activation proceeds down from higher levels to lower ones (Bar et al., 2006; Di Lollo, 2012; Di Lollo, Enns, & Rensink, 2000; Herzog & Manassi, 2015; Hochstein & Ahissar, 2002; Kahan & Enns, 2014; Lamme, 2006; Tapia & Beck, 2014).[2] Feedback processing is involved in identification, essentially the activation of stored semantic information that is incorporated in the perceptual object.[3] Processing of a stimulus prompts activation from LTM of information about the object it represents or resembles. The semantic features of the object have strong associative bonds with the surface features of the stimulus, so spreading activation from stored traces of the surface features (now integrated into a perceptual object) readily activates the semantic information. Thus, the end-product of perceptual processing is a perceptual product with semantic information integrated with surface sensory information. In effect the stimulus "king" activates the semantic representation of "king" in LTM by a process that involves spreading activation. These perceptual products or objects are entered into the EP, either directly or via sensory memory (Coltheart, 1980; Darwin, Turvey, & Crowder, 1972; Jacob, Breitmeyer, & Treviño, 2013; Ögmen, Ekiz, Huynh, Bedell, & Tripathy, 2013; Sperling, 1960).

Spreading activation models information storage in LTM as comprising nodes (representing individual items of information) connected by links to make a network (Collins & Loftus, 1975). Activation of a node results in activity spreading in parallel most strongly to adjacent nodes and continuing beyond that, gradually attenuating. The links can be assumed to represent associative bonds between nodes. Thus, when the stimulus "king" is presented, initially activation spreads to the node or nodes comprising the semantics of "king", which are strongly associated with the surface features of the stimulus. Spreading activation continues beyond that and activates, to some degree, other traces in LTM that have strong associative bonds with the "king" trace. This might

---

[2] Low level processing in vision is generally regarded as fast, automatic, non-attentive, and local (Holcombe, 2009). It is concerned with analysis of features such as lines, edges, and angles. High level processing tends to be slower, to involve attentive processing, to focus on a more global analysis (such as, at the level of individuated perceptual objects rather than local feature detection), and to involve pre-existing structures, semantic processing, and featural continuity analysis (Holcombe, 2009). Instead of a linear feedforward sweep of processing from low level through to high level processing, recent research has shown that an early feedforward sweep from low to high level triggers a feedback sweep from high to low level, which may represent integration of local feature information with more global and semantic information to construct individuated, identified, and persisting perceptual objects.

[3] Semantics concerns meaning. In language it is distinguished from syntax, which concerns arrangements of words and phrases in grammatical structures (e.g. Branigan & Pickering, 2017); in perception and memory it is distinguised from acoustical and surface visual information (e.g. Baddeley, 2000; Coltheart, 1980). In perception, an example of semantic information would be object identification or categorization, such as perceiving the stimulus "A" as a letter of the alphabet rather than just a visual shape. An auditory tone can be perceived as the sound of a flute (or a bird), as having expressive significance through its relationship with other notes in a piece of music, as being an alarm call, and many other things. All of those involve activation of information stored in LTM and incorporation of that information into the percept by means of feedback or re-entrant processing.

include semantic or categorical associates such as “queen” and phonological associates such as “thing” (Pickering & Gambi, 2018). The likelihood and degree of activation of a given trace depends, in part, on the strength of its associative bond with the trace corresponding to the stimulus.

Spreading activation was originally proposed as a theory of semantic memory and individual nodes in memory were regarded as concepts, such as the concept of “king” (Collins & Loftus, 1975). But the basic idea of spreading activation need not be tied to a concept-based account of semantic memory. Any kind of information in LTM may be treated in terms of its associative bonds with other nodes or elements, and that includes hierarchical structure information, temporal information, and so on. Thus, if a listener is familiar with Beethoven's fifth symphony then there will be strong associative bonds in LTM representing the sequence of chords and their temporal hierarchical structure. When the listener hears the opening chord, the perceptual representation of that may be sufficient to activate the chords that follow the opening chord and their temporal hierarchical structure by spreading activation. Integrated representations in the EP are not created from nothing. They are created as the sequence of chords successively activates structural information in LTM that has the strongest associative bonds with the stimulus information. Previous experience with music leaves traces in LTM that represent tonality, musical phrase constructions, metrical and rhythmic structure, and other features of music, and these form the basis for the representation of the music in the EP (Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Tillmann, 2012). Spreading activation may supply any kind of information to the EP, to the extent that the information is associated with the input stimuli in LTM. Importantly, it supplies thematic connections between successive perceptual products, which are in effect temporal associative bonds between them.

Pickering and Gambi (2018) regarded spreading activation as a form of prediction. They reviewed studies showing, for example, that incomplete sentences tend to prompt eye movements to an object, the word for which would be a plausible (and hence predicted) continuation of the sentence (Kamide, Altmann, & Haywood, 2003). Presentation of “the” as a stimulus leads to activation of LTM representations with similar features; these representations tend to be strongly associated with the concept of the definite article, and that in turn is strongly associated with the concept of a noun as a temporal follower to a definite article. Thus, after a single word presentation, there is already a hypothetical thematic connection, if not yet a hierarchical structure, that functions as a prediction for the kind of word that will be presented next (Pickering & Gambi, 2018). Presentation of more words results in further iterations of the same process and generation of an increasingly elaborated and specified thematically connected hierarchical structure that includes projected future elements. As an example of supporting evidence, Bresnan and Ford (2010) took advantage of a corpus showing differences in relative frequency of use of alternative constructions in different English-speaking countries. Participants tended to predict the one of two alternative syntactic paraphrases that was more frequently encountered in their linguistic environment. This result is consistent with spreading activation as a mechanism for construction of hierarchical structure, because more frequently encountered continuations of a sentence are likely to be more strongly activated than less frequently encountered versions are, reflecting relative strength of the associative bonds. But it is not clear that predictions will always be generated. Prediction should perhaps be regarded as a common rather than a necessary feature of hierarchical structure construction; indeed, prediction is often neither feasible nor useful (Jackendoff, 2007). The primary function of the EP is comprehension; making sense of what is going on now by integrating it into a context of recent information. Whether or not prediction is involved, locating present input in a temporal hierarchical structure is an important part of comprehension.

Although strength of associative bonds tends to be a stable residue of experience, the likelihood of activation of a given link can be transiently affected by recent experience. In a priming manipulation, prior exposure to stimuli including a particular semantic element transiently increases the relative likelihood of activation of that element, with the consequence that it may be favoured as the interpretation of a subsequent ambiguous stimulus. Priming effects are consistent with the hypothesis of spreading activation (e.g. Meyer & Schvaneveldt, 1971). There is evidence that hierarchical structure construction is affected by priming manipulations, both within and between domains (Bigand, Madurell, Tillmann, & Pineau, 1999; Scheepers et al., 2011; Tillmann, 2012; Van de Cavey & Hartsuiker, 2016). Because priming effects are consistent with spreading activation, that evidence fits with the hypothesis that hierarchical structure is constructed in similar ways to other semantic information, involving matching to LTM traces and spreading activation.

Spreading activation is in effect a form of parallel processing and could lead to a proliferation of activated information that could overwhelm the capacity of the storage space. Such proliferation does seem to happen for semantic information but only on a very short time scale. Swinney (1979) and Tanenhaus, Leiman, and Seidenberg (1979) found evidence that all meanings of an ambiguous word are activated by its presentation, consistent with what spreading activation would predict, but Tanenhaus et al. (1979) found that meanings that do not fit the syntactic context are no longer activated after about 200 ms. They argued that meanings that turn out to be inappropriate are actively suppressed. The speed with which this happens indicates that it is an automatic process, and the short time scale indicates that the initial proliferation of activated meanings does not persist for long enough to impose much demand on WM capacity.

Suppressing inappropriate meanings for a word presented in a syntactic context may be rapid and automatic, but it is not clear that the same would occur for a developing syntactic structure. Two hundred milliseconds is close to the average duration of individual phonemes (Hickok & Poeppel, 2007). Resolving syntactic ambiguity on the time scale of typical sentences would therefore involve a time scale rather longer than 200 ms. In that case multiple activated interpretations might impose high demands on memory space. One possibility is that just a single syntactic structure is activated at a given time, a possibility exemplified by serial accounts of sentence comprehension such as the garden path theory (Frazier & Rayner, 1982). In serial accounts, a parsing mechanism adopts and runs with a single syntactic model of a sentence at an early stage, with the possibility of reinterpretation if the structure is disconfirmed by subsequent information. The alternative is parallel accounts in which multiple interpretations are activated simultaneously. The extreme possibility of a parallel parser that always maintains all syntactical constructions of a sentence has been disconfirmed (Frazier & Rayner, 1982; Gibson & Pearlmutter, 2000; Lewis, 2000) but there could be more limited versions of parallel models. For example, Gibson (1991) proposed a parallel model in which the number of possible interpretations activated is limited by storage resources. Recent evidence favours such limited parallel models over serial models, again consistent with spreading activation (Chen, Gibson, & Wolf, 2005; Dillon, Andrews, Rotello, & Wagers, 2019). It is likely, then, that the number of models of hierarchical structure active at a given moment is limited by processing and storage costs. In that case, the most strongly activated model is favoured.

Gibson (1998) proposed that, in the case of language processing, there are two kinds of processing costs. One is memory cost which is to do with the resources available to store the words or partial sentence presented so far. The other is integration cost, the resources required to integrate new input into the existing structure. There is evidence supporting this distinction (Felser, Clahsen, & Münte, 2003). Memory cost can be substantial for the EP but there are ways of dealing with that. Some information in the EP may be transferred to LTM, from where it is readily accessible because of strong associative bonds with information currently in the EP (Norris, 2017; Oberauer, 2002). The integration cost in the domain-general account proposed here, however, is minimal because matching and spreading activation are automatic processes that take up little if any of WM processing capacity.

Different kinds of information are associated with each other in LTM representations to the extent that those different kinds of information are processed into a coherent, connected structure of information in the EP, and to the extent that that information is transferred to LTM. Hierarchical structure information may facilitate retrieval of associated semantic, prosodic, and timing information by spreading activation; in fact any kind may facilitate retrieval of any other kind depending on the strength of association between their representations in LTM. So hierarchical structure processing is not separate from other kinds of information; all are connected and integrated in the construction process.

In summary, construction of the EP is a rapid automatic ongoing process involving matching of current perceptual input to representations of information in LTM, and spreading activation from those representations to other representations with which they have the strongest associative bonds. Representations are activated in parallel but the most strongly activated set of representations tends to be adopted as the current interpretation and, often, predictor of the next input. That is subject to continual modification as further perceptual information enters. As an information structure develops in the EP, so (some) information in it is likely to be transferred to LTM from where it is readily accessed because of its strong associative bonds with information retained in the EP. All of this applies to all the kinds of information in the EP, so that domain-specific kinds of information are integrated with hierarchical structure information. Within that general account, many of the details are determined by domain-specific processing. For example, in the case of language, spreading activation may be involved in the generation of syntactic structures as discussed above, but domain-specific WM processes may determine which of the activated syntactic constructions is adopted as the interpretation of the sentence in question.

2.2. *Evidence for hierarchical structure representation in multiple domains*

If hierarchical structure is a key domain-general feature of information in the EP, it should be possible to find evidence for hierarchical structure processing in multiple domains, and specifically evidence for representation of hierarchical structure information in WM. There is relevant evidence from the domains of music processing, language, action perception and execution, and arithmetic, and this will be briefly reviewed.

2.2.1. *Music*

"Like language, music is a human universal in which perceptually discrete elements are organized into hierarchically structured sequences according to syntactic principles" (Patel, 2003, p. 674). Evidence for the universality of music and some of its basic features across human cultures has been reported by Mehr et al. (2019). Music can be formally analysed in terms of syntactic or other kinds of hierarchical structures, although there are conflicting accounts of how this should be done (Bigand et al., 1999; Koelsch et al., 2013; Krumhansl, 1990; Lerdahl, 2001, 2015; Lerdahl & Jackendoff, 1983; Narmour, 1983; Patel, 2003; Rohrmeier, 2011), but the issue of interest in this sub-section is whether information about music in WM includes hierarchical structure information.

Several studies have shown that music is processed in terms of hierarchical structure (Bigand, 1997; Bigand et al., 1999; Collins, Tillmann, Barrett, Delbé, & Janata, 2014; Cuddy, Cohen, & Mewhort, 1981; Zhang, Jiang, Zhou, & Yang, 2016). Evidence for representation of recursive structure in musical sequences was found by Martins, Gingras, Puig-Waldmueller, and Fitch (2017). There is evidence for detection of nested relations in sequences of tone stimuli by infants aged five months (Winkler, Mueller, Friederici, & Männel, 2018). Nesting or embedding is a common feature of hierarchical structure; the results do not show any other form of hierarchical structure representation, but nesting could be a developmental precursor to more complex hierarchical representations. Those studies demonstrate hierarchical structure processing in the domain of music, but they do not suffice to show whether and how hierarchical structure information is held in WM. That is the kind of evidence that would be needed to support the hypothesis of the EP in the domain of music.

2.2.1.1. *Evidence for hierarchical structure information in WM for music.*

Several studies have found evidence that judgments of a final chord in a series vary depending on chords presented earlier in the series (e.g. Bigand & Pineau, 1997), but that alone is not sufficient to establish that the judgment was affected by a change in hierarchical structure rather than some other feature. Specific manipulations of hierarchical structure are required for that. Two studies have found evidence for that.

Bigand et al. (1999) manipulated the chord sequence preceding a final chord at three hierarchical levels of structure with sequences of 14 chords. They found significant effects on responses to the target chord at the end of the sequence, even when the last six chords prior to the target chord were identical. This implies representation of hierarchical structure in memory over more than seven chords.

In a similar vein, Koelsch et al. (2013) made use of two Bach chorales both having an ABA structure ending with resolution on the tonic over a span of (in one case) sixteen chords. Koelsch et al. analysed the chorales as having a hierarchical tree structure with three levels. They altered the first half of each chorale by transposing the music up or down by a fixed number of intervals, in such a way that the final chord no longer brought closure to the structure, even though it completed a locally correct cadence. They found evidence for differences in ERPs evoked by the final chord depending on whether the original or modified chorale was presented. Since the second half of the chorale was the same in both versions, and the local structure of the second half provided closure, the results show a response to a non-local structural variation on the scale of the sixteen chords of the chorale. Collins et al. (2014) argued that studies of hierarchical structure representation have confused syntactic structure with probabilistic (non-hierarchical) relations between successive notes. If that is true of earlier studies, it cannot be true of the study by Koelsch et al. (2013), and in fact Koelsch et al. demonstrated that models based on probabilistic relations between adjacent notes or chords could not account for their results. Their study therefore provides the strongest evidence for the representation of hierarchical structure in music in WM.

There are two major problems with evidence of the kind reviewed here. One is that it is not clear exactly how hierarchical structure information is represented in memory. The other is whether there are other ways of interpreting the evidence. These will now be considered.

2.2.1.2. *How is hierarchical information in music represented in WM?.*

A formal analysis of multi-level hierarchical structure might predict the results of the studies, but that is not proof that the full hierarchical structure is represented in the surviving information structure. Syntactic properties of stimuli may be transformed into some other kind of representation (Lerdahl, 2015; Lerdahl & Krumhansl, 2007). An example of another kind of representation that is specific to music is the joint accent structure proposed by Jones (1987). According to Jones (1987), the joint accent structure is a way of structuring musical inputs in terms of temporal information, specifically rhythm and meter, and accents, which are tones that stand out from other tones on some auditory dimension; Ellis and Jones (2009) specified pitch, intensity, timbre, and duration. Accents are usually related to metrical structure and can be a way of indicating the metrical structure of the music. Music is therefore represented in terms of a structure defined by accent and rhythm. Some notes have features that lead to them being identified as accents and others do not. New musical input is perceived in the context of the structure established up to that point, and contributes to the further development of that structure. The theory is in part an attentional theory, in that temporal regularity in the developing musical structure tends to entrain periodicity in attention. That in turn facilitates comprehension. For present purposes, however, it is the two main determinants of musical structure construction, temporal information and accents, that are of central interest. Temporal information will be the topic of the next main section. For this section the main concern is with the auditory features that determine identification of accents. Pitch, intensity, and timbre are prosodic features, and the proposal implies that prosodic information may be a major determinant of structural representations of music.

The simplest characterisation of prosody in language is as the features of spoken utterances that are not shared by written utterances (Shattuck-Hufnagel & Turk, 1996). Corresponding to that, musical prosody would be those characteristics of musical performance that are not in the written score, or at least a score without expression marks. These characterisations conceal a world of complication and contention which there is insufficient space to explore in this paper. However, there is general agreement that prosody can be analysed as hierarchically organised, and several prosodic hierarchies have been proposed; Shattuck-Hufnagel and Turk (1996) provided a useful summary of proposals up to that date. Formal properties of prosodic hierarchies do not match those of syntactic hierarchies, though in practice there is often considerable overlap between prosodic and syntactic hierarchies for given utterances.

Prosodic features form an important part of musical representations as well (Heffner & Slevc, 2015; Palmer & Hutchins, 2006). That is, prosodic features of music are perceived and remembered: evidence for this includes research on pitch height (Schellenberg & Trehub, 2003), tempo (Levitin & Cook, 1996), and variations in musical expression (Palmer, Jungers, & Jusczyk, 2001). Palmer and Hutchins (2006) argued that prosodic cues in music have functions associated with segmentation. For example, performers often lengthen tones as they approach a structural boundary in music, even though the notes are notated as all having the same length: “The patterns of tempo modulations often indicate a hierarchy of phrases, with the amount of slowing at a boundary corresponding to the depth of the phrase embedding” (Palmer & Hutchins, 2006, p. 254). Slowing could be regarded as an accent in a joint accent structure (Jones, 1987). Thus, there can be a hierarchical prosodic structure that, when in the service of segmentation, overlaps with the formal hierarchical structure of the music. It has been shown that prosodic features influence segmentation of music by listeners and also influence memory for musical passages (Palmer et al., 2001). This supports the hypothesis that hierarchical prosodic structure is a feature of WM representations of music. Heffner and Slevc (2015) made a case that there are close parallels between prosody in speech and in music: if there can be multi-level prosodic hierarchies in speech, there can be multi-level prosodic hierarchies in music too. Hierarchical prosodic structures in speech are often closely associated with hierarchical syntactic structures, not least because they contribute to segmentation. That would imply that hierarchical prosodic structures in music also tend to be closely associated with hierarchical musical structures, at least to the extent that they subserve structure-related functions such as segmentation and accentuation, as Heffner and Slevc (2015) argued.

It is likely, therefore, that the information structure of music integrates multiple kinds of information into a unified hierarchical representation of the time course of a short stretch of music. The amount of information involved in such a representation appears greater than the likely capacity of WM, particularly if the representation can extend over 16 chords (Koelsch et al., 2013) with associated syntactic structure information and structured tonal and prosodic information concerning multiple perceptual features. There is evidence that WM capacity for non-verbal auditory stimuli may be limited to one or two tones (Li, Cowan, & Saults, 2013; Prosser, 1995). However, capacity limitations in WM can be compensated by accessible storage in LTM (Mallett & Lewis-Peacock, 2018; Norris, 2017; Oberauer, 2002), and that could contribute to the representation of hierarchical structure in music in studies such as that by Koelsch et al. (2013). There is evidence that melodies (specifically pitch and temporal relations between notes) are stored in LTM after just two exposures, with accurate recall even a week later (Schellenberg & Habashi, 2015). The study did not test whether the hierarchical structure information in the melodies was also stored in LTM, but it could be reconstructed from the recalled music just as it could have been constructed from exposure to the original stimuli. So LTM can certainly compensate for limited capacity WM storage.

Consistent with the present proposal that spreading activation is involved in the construction of hierarchical structures, Bigand et al. (1999) proposed a spreading activation model according to which a short-term auditory context primes expectations by activating knowledge in LTM through associative connections. This can include both specific (episodic) memories of previously heard musical phrases or pieces and acquired knowledge of general features of musical structure. Supportive evidence comes from a study of music perception by Bigand and Pineau (1997). Thus, the information structure for a passage of music combines information stored in WM with information accessed from LTM and there is a continual interplay between perceptual information entering the EP and spreading activation eliciting stored representations of musical information of multiple kinds.

2.2.1.3. Alternative hypotheses. Schellenberg and Habashi (2015) found that the pitch and duration information that defines melodic phrases can be stored in and accurately recalled from LTM after only two exposures. This raises the possibility that responses in studies of hierarchical structure processing could be based just on memory for melodies (or chord sequences) or other surface features of musical stimuli with no hierarchical structure processing. There are two problems with this alternative hypothesis. One is that Schellenberg and Habashi (2015) did not test whether any hierarchical structure processing occurred. Thus, it is possible that hierarchical structure information was stored in LTM along with, and integrated with, the melodic information. Second, in the study by Koelsch et al. (2013), both electrophysiological and behavioral responses differed depending on whether the final chord closed a hierarchical structure or not. If only melodic information was processed and stored, the final chord would just be another event tacked on to the chord sequence, another note in a melody. Hearing a melodic phrase or chord sequence as having come to a final point means that at least some of its hierarchical structure has been processed; without that, there would be no closure to melodic phrases. In the Koelsch et al. (2013) study the dependency was nonlocal; that is, the sense of closure (or lack of it) depended on events occurring many chords earlier. If only melodic information was stored, perception of the final chord would not be affected by events much earlier in the sequence.

A second alternative hypothesis is that musical structure is represented in WM with little or no dependence on information stored in LTM; that is, stored representations of hierarchical structure would not be activated. Leman (2000) and Bigand, Delbé, Poulin-Charronnat, Leman, and Tillmann (2014) proposed that listeners evaluate a current auditory input (note or chord) by the extent to which it resembles or fits with a tonal context comprising the recent history of the pitches of notes or chords as a decaying representation in auditory WM (AWM). They argued that there need be no hierarchical structure processing, and indeed no representation of tonal hierarchy information in LTM, just a judgment of similarity, meaning degree of fit with the tonal context as represented in AWM (Bigand et al., 2014; Leman, 2000). AWM is assumed to have limited capacity and rapid decay: Leman (2000) posited a value of 1.5 s for half decay on the basis that this gave a good fit with data from a study by Krumhansl and Kessler (1982). The half-decay time was defined as the time in which an input decayed to half its initial amplitude. If decay is a linear function of time, that would give a full decay time of about 3 s, which is similar to that found for the phonological loop (Baddeley & Hitch, 1974). An exponential decay curve would yield a longer decay time. It is likely, therefore, that the “echoic memory”, the term used by both Leman (2000) and Bigand et al. (2014), is in fact a WM store (Caclin & Tillmann, 2018; Pechmann & Mohr, 1992; Schendel & Palmer, 2007; Schulze & Koelsch, 2012). Bigand et al. (2014) tested their model's capacity to account for the results of several studies with long context dependency (e.g. eight chords in Bigand & Pineau, 1997). They found a good fit for most of the studies but were not able to model results found by Tekman and Bharucha (1998) and Bigand, Poulin, Tillmann, Madurell, and D'Adamo (2003).

Despite the success of the model on other datasets, the two failures are of critical importance. Tekman and Bharucha (1998) set up a discriminative test of psychoacoustic similarity and implicit knowledge in the form of music conventions, as possible sources of expectations for chord continuations. They found that, with stimulus onset asynchrony >500 ms, expectations were driven by implicit knowledge, which would be held in LTM. This result counts strongly against similarity-based models on the time scale of WM. Bigand et al. (2003) argued that their findings showed the influence of knowledge of Western music rules in perception and judgment, which would be held in LTM. That would contradict the predictions of models such as that by Leman (2000) in which judgment is based on similarity between the presented chord and the decaying WM trace of the pitches of the previous chords. Also, Bigand et al. (2014) did not test their model against data from the study by Koelsch et al. (2013). In that study judgment of a final chord depended on chords presented up to 16 chords previously. That would be beyond the capacity of AWM. Bigand et al. (2014) commented that the failures of the model set in the context of its many successes could be addressed by modifications to the model rather than rejecting it. That could be the case and the matter is far from settled. On the other hand, in common with other computational models, the AWM model has free parameters that are set empirically to maximise fit to the data, and that can make it difficult to specify conditions for falsification. It is also relevant to note that temporal order information is critical to the similarity-based accounts (see, e.g., Leman, 2000, p. 505), but there is no explanation as to how temporal information is represented. Since the surviving information about events at different times is all there in perception and WM at one time, this is clearly important.

There is a considerable overlap between representations of passages of music in terms of tonality and representations in terms of a temporal hierarchical structure: the temporal hierarchical structure would make no sense without some understanding of tonality and the places of different notes in the tonal scale. Equally, understanding of tonal relations is presumably based to some extent on experience, which would imply that knowledge of tonality is in, and activated from, LTM, when music is being listened to. Thus, even if perception of music was based on degree of fit with the tonal context information in WM, activated knowledge from LTM would still play a significant role in constructing that representation. And hierarchical structure processing accounts do not deny the decay of information in WM. So, at present, the similarity hypothesis has some supporting evidence but has problems accounting for evidence that implicit knowledge activated from LTM is involved in music perception (Bigand et al., 2003; Tekman & Bharucha, 1998). The case for hierarchical structure information being activated and held in WM representations of music, therefore, remains strong.

2.2.1.4. Loss of information beyond the boundary of the EP. If the EP is a distinct information structure, then memory for past information, that is, outside the temporal boundary of the EP, should show different properties. Consistent with that, research on longer passages of music has shown little if any retention of structural features on longer time scales. Cook (1987) presented passages of music with durations ranging from 30 s to 6 min, in which the tonality at the end was different from the main tonality of the passage. This should have disrupted the sense of closure or completeness of the piece of music, but Cook found a significant effect on ratings of coherence and completion only for the shortest piece, and no significant effects on ratings of pleasure or expression. Other studies have re-ordered pieces of music on long time scales, supposedly disrupting the global organization present in the original, but found no significant effects on various ratings related to aesthetic qualities (Gottlieb and Konecni, 1985; Karno & Konecni, 1992; Konecni, 1984; Lalitte & Bigand, 2006; Tillmann & Bigand, 1996, 2004; Tillmann, Bigand, & Madurell, 1998).

In summary, perception of music is accompanied by integration of information about the recent history of the music into a temporal hierarchical structure over a span that can extend up to at least 16 chords into the past. Information about the more temporally distant past of the music is much more tenuous and may consist of relatively abstract semantic summaries that represent some information about repetition, tension and resolution, and general features of the development of the music over time. The research therefore supports the hypothesis of a qualitative boundary between the EP and retained information about the more distant past.

2.2.2. *Language*

It is well known that most linguistic utterances have hierarchical syntactic structure. It is also well established that parsing of syntactic structure is incremental, with development and adjustment as successive components of the utterance are perceived (Just & Carpenter, 1980; Tyler & Marslen-Wilson, 1977). Regardless of how parsing is done (see, e.g., Bock, Loebell, & Morey, 1992), what matters for present purposes is that it takes time. Pellegrino, Coupé, and Marsico (2011) found an average speech rate varying between 5.2 and 7.8 syllables per s across seven languages, and this finding was replicated by Coupé, Oh, Dediu, and Pellegrino (2019) for seventeen languages representing nine language families. The time scale of words ranges from ~200 ms to over 1000 ms (Herzog, Kammer, & Scharnowski, 2016; Hickok & Poeppel, 2007). This means that sentences are produced on a time scale of seconds. Contrasting with that, the human auditory system has remarkably fine temporal discrimination. Specifically on processing of speech stimuli, there is evidence for a temporal window of analysis of about 20–50 ms that is critical for determining the temporal order of the segments that make up a syllable or word (Chait, Greenberg, Arai, Simon, & Poeppel, 2015; Hickok & Poeppel, 2007; Telkemeyer et al., 2009); Hickok and Poeppel (2007) gave the example of distinguishing between “pets” and “pest”.

All of that means that speech perception involves processing large amounts of information on time scales ranging from a few milliseconds to several seconds. Much of speech perception must be accomplished by automatic processes with large capacity (Kljajevic, 2010; Patel, 2003). Parsing of syntactic structure probably falls into that category. The typical time scale of a sentence takes it well beyond the boundaries of temporal integration in perception. Large amounts of information can be stored in auditory sensory memory (Darwin et al., 1972) but, while sensory memory can store information such as perceptual features and objects, it has not been established that it can store syntactic information; furthermore, information in auditory sensory memory decays to the limited capacity of WM in about 1 s (Darwin et al., 1972). Ambiguous syntactic structures require early components to be held in some kind of store until a later event in the sentence resolves the interpretation, which implies a time scale longer than that of auditory sensory memory. Processing such sentences requires the computational resources of WM, probably supplemented by LTM (Gibson & Hickok, 1993; Kljajevic, 2010; Pickering & Barry, 1991). It is likely, therefore, that syntactic structure information must be housed in WM before processing of most sentences is complete. Indeed, analyses of difficulty in processing linguistic input seem to depend on the proposition that capacity is limited, despite the fact that much of the processing is automatic (Levy, 2008), and the primary capacity limitation is likely to be on the amount of information about the input sentence that can be held in an active state in WM.

It is not controversial that syntactic structures, both perceived and produced, are often if not always hierarchical (Pickering & Gambi, 2018). Sentences often involve multiple hierarchical levels and sentence structures often exhibit embedding and recursion (Fitch & Martins, 2014; Jackendoff, 2007). In the model of WM proposed by Baddeley and Hitch (1974), the phonological loop was proposed as a store of acoustic and articulatory information (Baddeley & Hitch, 1974, 2019) and was not proposed as a store of syntactic information (see also Caplan & Waters, 2013, for a history of research on that issue). But there is evidence for a WM store of syntactic information that is separate from the phonological loop (Jackendoff, 2007). There have been reports of patients with normal syntactic comprehension despite having deficits in the phonological loop (Martin, 1987; Martin & Romani, 1994; Vallar & Baddeley, 1984). Martin and Romani (1994) argued that their results supported the hypothesis of separate WM stores for phonological, semantic, and syntactic information. The syntactic store would be capable of retaining hierarchical structure information on the scale of sentences in normal discourse. Martin and Romani (1994) argued for an interactive activation model in which the evolving representation of a sentence represents connections or associations between phonological, semantic, and syntactic features. That again is consistent with the spreading activation hypothesis.

Sentences have phonological structure but it is independent of syntactic structure and has different units and combination rules (Jackendoff, 2007; Liberman & Prince, 1977). Yet, as Jackendoff (2007) pointed out, phonological structure tends to be correlated with syntactic structure and is also to some degree hierarchical. In this respect the relationship between syntax and phonological structure in language appears similar to that between hierarchical structure and prosodic structure in music, as discussed in the previous section. As an example, there is a relationship between the length of pauses between intonation units and the closeness of syntactic relationships (Frazier, Carlson, & Clifton, 2006), and this is paralleled by a relationship between pause length between musical phrases and the closeness of their relationships in hierarchical structures of music (Palmer & Hutchins, 2006). Frazier et al. (2006) argued that prosodic structure facilitates the retention of auditory stimuli in memory. They gave the example of nursery rhymes, where highly constrained and predictable prosodic patterns facilitate retention for inexperienced listeners. Indeed, nursery rhymes sit near the border between speech and music: it is likely that constrained and predictable prosodic patterns facilitate memory for music as well as for language. Thus, in both language and music, hierarchical (syntactic) structure processing and prosodic processing are independent but also to some degree integrated, and the integration between them facilitates comprehension and memory.

Jackendoff (2007) also reviewed evidence that semantic processing is independent of syntactic processing. Semantic structure is correlated with syntactic structure and the two are integrated in full linguistic representations. Thus, when a sentence has been processed, an integrated representation of semantic and syntactic and phonological or prosodic features is present in WM. Under the present hypothesis, this integrated representation would form (part of) the content of the EP. But what happens after that, when the complete sentence has been processed? As in the case of music, if the EP is a distinct informational construct, information about sentences previous to the one currently represented in the EP should be significantly attenuated.

Lombardi and Potter (1992) argued that, when the end of a sentence is reached, a conceptual representation of the sentence is retained in WM along with the verb, but not with the syntactic information. They argued that the verb and the conceptual representation suffice to regenerate the syntactic structure in recall, by activating associated structures in LTM; this is the regeneration hypothesis. In addition, recently activated structures are transiently more accessible than structures that have not been activated recently. These processes support a high level of accuracy in short-term recall of sentences. As evidence, after being given a sentence to recall, participants were set another task involving presentation of a list of words. One of the words would be a synonym of a verb in the target sentence, but of a sort that would imply a different syntactic structure. Participants who incorrectly recalled the synonym as having been in the target sentence tended to reconstruct the syntax of the sentence to fit with the incorrect verb. This implies that the original syntactic structure was not retained in WM, but was reconstructed from (in this case, inaccurate) cues to the key verb and the semantics.

The regeneration hypothesis predicts that, when two syntactic forms are both compatible with the verb and the conceptual representation, one of which was the originally presented form, either form may be generated in recall. In fact, the originally presented form was regenerated on the majority of occasions. Potter and Lombardi (1998) found support for the hypothesis that the presented syntax exerts a priming effect, essentially increasing the transient accessibility of the presented structure from LTM relative to the unpresented structure. Thus, their research is consistent with the hypothesis that syntactic structure of sentences previous to the current one is not retained in WM, but that it can be accurately regenerated on short time scales from available information about semantics, the verb in the sentence, and the priming effect of the presented sentence. The syntactic comprehension results found by Martin and Romani (1994) could then be interpreted as resulting from the operation of this regeneration mechanism. Syntactic information may not be lost completely. There is evidence for significant memory for syntactic information after presentation of prose passages greatly exceeding WM capacity (Gurevich, Johnson, & Goldberg, 2010; Sachs, 1967), but Sachs (1967) found markedly superior recognition of semantic changes than syntactic ones, indicating that what is stored in LTM is predominantly semantic information.

At a given moment, therefore, there is a comparatively rich representation in the EP of an ongoing sentence with semantic, syntactic, prosodic, and phonological features. Information about sentences previous to the one in the EP can persist in LTM but is much reduced and comprises mainly semantic information and cues to regeneration of sentence structure. In terms of the present proposal, therefore, the boundary of the EP is marked by selective loss of information, particularly about hierarchical structure, resulting from the removal of syntactic and perhaps other information from the representation (Lewis-Peacock, Kessler, & Oberauer, 2018). That would be consistent with the evidence from research on music discussed earlier that hierarchical structure information is lost from the EP after an event boundary (such as the resolution of a musical phrase onto the tonic), and more distant past musical information is predominantly relatively abstract semantic summaries sitting outside the EP, as discussed in the sub-section on music above.

2.2.3. *Generation and perception of action*

It is well established that goal-directed actions can be analysed in terms of their hierarchical structure (Byrne & Russon, 1998; Dawkins, 1976; Fitch & Martins, 2014; Lashley, 1951). Accomplishing a goal such as making a cup of coffee can be represented as a hierarchically organised set of sub-goals and associated actions (Jackendoff, 2009). For example, a sub-goal of making coffee may be getting water, and getting water may involve grasping a kettle, positioning it under a tap, and turning the tap on. Thus, a temporal sequence of individual actions has a hierarchical structure reflecting the organised sub-goals involved in achieving the higher-order goal of the action. The questions for this section are, is action generation represented in the brain as hierarchically organised, are actions perceived in terms of hierarchical structure, and how do these things relate to the EP?

Goal-directed action generation involves hierarchical structure, in the sense that accomplishing a higher-order goal may require the setting of sub-goals and executing individual actions in the correct order to meet both the sub-goals and the overall goal. Lashley (1951) argued that execution of such action sequences requires the higher-order goal to be held in a temporary store while sub-goals are being achieved. For example, it is important to remember that the kettle is being filled with water in order to contribute to the higher-order goal of making coffee. Information must be stored on the time scale of the whole action sequence. As Fitch and Martins (2014) pointed out, “the more complex and multileveled the hierarchy, the greater the capacity of the intermediate storage mechanism must be" (p. 95). As we have seen, storage capacity limitations in WM can be compensated by assigning information to LTM where it has transiently increased accessibility. The components of familiar action sequences have strong associative bonds that further increase their accessibility.

Hierarchically structured goal-directed action begins to develop in the second year of life and soon develops to the point where recursive loops are introduced for error correction (Connolly & Dalgleish, 1989). Several studies by Greenfield and colleagues (e.g. Goodson & Greenfield, 1975; Greenfield, 1976; Greenfield, Nelson, & Saltzman, 1972) have documented the development of increasing hierarchical complexity in action sequences in early childhood.

There is also evidence of representation of hierarchical structure in action planning and generation in the brain (Balaguer, Spiers, Hassabis, & Summerfield, 2016). Koechlin and Jubault (2006) compared hierarchically structured action sequences with sequential actions lacking in hierarchical structure. They found that the hierarchical structure of the action sequence was reflected in a neuroanatomically ordered arrangement of regions in Broca's area, with a progression from lower to higher hierarchical levels running from posterior to anterior regions: "anterior [Broca's area] regions show phasic activation at boundaries of superordinate chunks only, providing evidence that these regions are specifically involved in selecting or inhibiting superordinate action chunks" (p. 968). Martins, Bianco, Sammler, and Villringer (2019) found that actions generated using a recursive hierarchical embedding rule specifically activated a network of several areas shown in previous research to be involved in motor learning, planning, and imagery while actions generated not using that kind of rule did not do so. It is likely that those results identify a process of transforming the hierarchical structure rules into action generation plans. There is clearly much still to learn about brain processes involved in hierarchically structured action generation, but the evidence does at least indicate that there are specialised areas for dealing with hierarchical structure in actions, particularly Broca's area.

The EP should be involved in both generation and perception of action. That is, perception of action should involve the construction of a hierarchical structure of the ongoing action in the EP. The representation that develops while perceiving another person making a cup of coffee should be similar, in terms of hierarchical structure, to the representation that develops while planning the making of a cup of coffee, other things being equal. In general, there is abundant evidence for close links between action perception and generation. For example, according to the theory of event coding (Hommel, Müsseler, Aschersleben, & Prinz, 2001), a common representational medium encodes and stores both perception of action and the representations involved in action planning. According to Hommel (2015), "the basic units of both perception and action… are *sensorimotor* entities, in the sense that they are activated by sensory input (=perception) and controlling motor output (=action)" (p. 2). There is much supportive evidence for this common coding hypothesis (e.g. Amer, Gozli, & Pratt, 2018; Elsner & Hommel, 2001; Hamilton, Wolpert, & Frith, 2004; Knoblich, Seigerschmidt, Flach, & Prinz, 2002; Prinz, 1997; Tipper, 2010; Tucker & Ellis, 1998, 2004). Given this, it would be expected that hierarchical structure in action generation would be represented in action perception, on the grounds of common coding of the two kinds of information. There is indeed evidence for that.

Maffongelli et al. (2015) presented sequences of still photographs representing simple actions such as making coffee which, as we have seen, can be analysed as having hierarchical structure. Participants viewed either a canonical sequence, a sequence in which the order of two photographs was inverted, resulting in a violation of the hierarchical structure, or a canonical sequence in which an inappropriate object was used, but wth no violation of hierarchical structure. They found ERP responses differentiating between the structure violation and the other stimuli. Maffongelli et al. (2018) ran a similar experiment with simpler stimuli and with participants aged about six months, and obtained similar results. This is consonant with previous studies showing that infants aged about six months can represent actions in terms of action goals (e.g. Daum et al., 2009; Falck-Ytter et al., 2006; Luo, 2010). Interestingly, it suggests that hierarchical structure processing may be operative even in prelinguistic infants. These are isolated studies but, as we shall see in the next sub-section but one, there are other studies that have argued for a common neural representation in multiple domains, and the representation of action is one of those.

2.2.4. A brief note on mathematics

Exact mathematics has hierarchical structures that formally resemble those found in the domains considered in the preceding subsections. The term "exact" is used because research in mathematical cognition supports a division between approximate arithmetical processing and exact calculation (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Pica, Lemer, Izard, & Dehaene, 2004). In approximation, the evidence suggests an immediate sense of numerosity for numbers up to four or five and approximate estimation for making arithmetical judgments about larger numbers. Exact calculation appears to be a separate system and to be linked to language (Dehaene et al., 1999). In exact calculation, Scheepers et al. (2011) pointed out that "Mathematical equations possess hierarchical structures that may resemble the syntactic structures of linguistic expressions" (p. 1320). The evidence for a close relationship between exact calculation and language (Dehaene et al., 1999) would be consistent with that assertion. Scheepers et al. (2011) gave the following examples:

"Example 1a: 80 - (9 + 1) x 5
Example 1b: 80 - 9 + 1 x 5" (p. 1320).

Although the two equations present the same numbers and arithmetic operation symbols in the same order, differing only in the presence or absence of parentheses, they are not solved in the same way. Example 1a has a hierarchical structure where one operation is nested inside others. Scheepers et al. argued that the structural difference between the two equations is analogous to the structural difference between high and low attachment of relative clauses in the example ambiguous sentence "I visited a friend of a colleague who lived in Spain". In one interpretation of this sentence there is a nested phrase: "I visited a friend (of a colleague) who lived in Spain". This relationship between calculation and language suggests that hierarchical structure in exact calculation may also be entered into the EP. However there does not appear to have been any research on the representation of hierarchically structured mathematical operations in WM. There have been studies supporting the hypothesis of a domain-general representation for language and exact calculation, of which the study by Scheepers et al. (2011) is an example, but nothing analogous to the study of hierarchical structure processing in music by Koelsch et al. (2013) has yet been attempted. However, there is evidence for a common mechanism for hierarchical structure processing shared by language, music, action representations, and exact calculation (Van de Cavey & Hartsuiker, 2016).

2.3. Domain-general or domain-specific hierarchical structure processing?

There are obvious domain-specific features of action, language, music, and mathematics. These primarily concern fundamental units (e.g. words v. tones), semantic and (in the case of language and music) prosodic features (see Jackendoff, 2009, for more on the differences between language and music). Moreover, kinds of hierarchical structures may also differ between domains. For example, Asano and Boeckx (2015) argued that syntax in language, meaning the rules for combining words to make sentences, is specific to language and does not generalise to music. However, there are features of temporal hierarchical structure that are abstract or domain-general. These include the basic operations and rules that make hierarchical structure hierarchical: those discussed earlier include the Merge operation, temporal integration, embedding, and recursion. It could be argued that the fact that hierarchical structure processing occurs in multiple domains of cognition is not a coincidence, but reflects a common foundation for domain-specific processing in domain-general temporal hierarchical structure processing. Indeed, there have been proposals that there is a domain-general processing mechanism for hierarchical structure (Kljajevic, 2010; Koelsch, 2012; Patel, 2003, 2008).

Cutting a long story short, there is a considerable amount of evidence that supports the hypothesis of a domain-general processing mechanism for temporal hierarchical structure. This evidence is of two main kinds. One kind involves evidence for cross-domain structural priming effects or interference effects, indicating a shared processing resource for the domains in question (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Fiveash, McArthur, & Thompson, 2018; Fiveash, Thompson, Badcock, & McArthur, 2018; Hoch, Poulin-Charronat, & Tillmann, 2011; Pozniak, Hemforth, & Scheepers, 2018; Scheepers et al., 2011; Scheepers, Galkina, Shtyrov, & Myachykov, 2019; Scheepers & Sturt, 2014; Slevc, Rosenberg, & Patel, 2009; Steinbeis & Koelsch, 2008; Tillmann, 2012; Van de Cavey & Hartsuiker, 2016; Van de Cavey, Severens, & Hartsuiker, 2017; Zeng, Mao, & Liu, 2018). Van de Cavey and Hartsuiker (2016) found cross-domain priming effects covering the four domains discussed above: relative clause sentences, music, exact calculation, and structured descriptions of actions.

The second kind of evidence concerns the same brain area being involved in processing hierarchical structure in different domains. There is evidence that the inferior gyrus, and specifically Broca's area (BA 44 and 45), is involved in processing hierarchical (syntactic) structure in language, music, action plans, and exact calculation (Abrams et al., 2011; Chen, Wu, Fu, Kang, & Feng, 2019; Cheung, Meyer, Friederici, & Koelsch, 2018; Chiang et al., 2018; Koechlin & Jubault, 2006; Kunert, Willems, Casasanto, Patel, & Hagoort, 2015; Maess, Koelsch, Gunter, & Friederici, 2001; Makuuchi, Bahlmann, & Friederici, 2012; Nakai & Okanoya, 2018; Nakai & Sakai, 2014; Opitz & Friederici, 2007; Pallier, Devauchelle, & Dehaene, 2011; Schell, Zaccarella, & Friederici, 2017). On the other hand, the evidence has been criticised on methodological grounds (Fedorenko & Varley, 2016), and there is also evidence against the hypothesis of domain-general processing (Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012; Fedorenko & Varley, 2016; Pritchett, Hoeflin, Koldeweyn, Dechter, & Fedorenko, 2018; Varley, Klessinger, Romanowski, & Siegal, 2005). Against that, it has been argued that the results might show domain-specific capacities with different localisations, and do not refute the possibility of domain-general hierarchical structure processing (Scheepers et al., 2019). Indeed the evaluation by Fedorenko and Varley (2016) did not touch the evidence from cross-domain priming and interference studies (e.g. Van de Cavey & Hartsuiker, 2016).

Despite the large amount of evidence that is consistent with the hypothesis of a domain-general hierarchical structure processing mechanism, it is not confirmed beyond doubt at present, and is just one among a number of possibilities being debated in the literature (e.g. Bigand et al., 2014; Chiang et al., 2018; Friederici, 2019; Peretz, Vuvan, Lagrois, & Armony, 2015). Another possibility is that there are just multiple, domain-specific hierarchical structure processing mechanisms, and there are similarities in the way they operate that reflect general features of temporal hierarchies (Asano & Boeckx, 2015).

The EP is an information structure that could involve multiple WM stores depending on the kind or domain of information that is in it at any given moment. Because of that, it makes no difference in principle to the EP whether the hierarchical structure information in it is constructed solely by domain-specific processes or whether there is a domain-general component. Although the question of domain-general hierarchical structure processing remains open, the research reviewed in this section supports several propositions about the EP: the EP represents hierarchical structure information in several domains, including music, language, action planning and perception, and exact calculation; the extent of hierarchical structures in the EP is limited by capacity limitations supplemented by the possibility of transiently accessible storage in LTM; hierarchical structures of information in the EP are integrated with semantic and prosodic information, and indeed prosodic hierarchical structures could be represented there; the EP represents a boundary on hierarchical structure information, in that hierarchical structure information is weakly represented in or absent from memory for more temporally distant events, which are represented in terms of incomplete semantic summaries.

3. Temporal information

Perhaps the most important component of the EP is time marking information. Items of information in the EP are labelled with time information. If that information were lost or scrambled then the information structure itself would be scrambled, or would disintegrate into isolated items. The main aims of this section are to identify the kinds of time marking information in the EP and their functional roles, and to discuss the mechanisms by which the timing information is generated and applied to information in the EP.

It is proposed that at least two kinds of temporal information are represented in the EP: ordinal temporal information (hereafter just "ordinal information") and duration/interval information. Ordinal information is critical to the representation of hierarchical structure information: changing the information about the ordinal relations between the components changes the hierarchical structure, with consequent changes in meaning, and losing ordinal information altogether would mean losing the hierarchical structure altogether, resulting in a mere collection of isolated items of information. To illustrate, Tillmann and Bigand (2001) found that scrambling the order of chords in a short series had strong effects on reported musical coherence, a clear indication that temporal order information is preserved in short-term memory for music.

The importance of duration information may be variable, depending on what sort of information is being processed. It should be noted that "duration" here means the duration of a stimulus or an event a or series of stimuli/events and not just the passage of time itself. In music, duration information is important because it permits the representation of notes or chords of different durations, as well as metrical structure and tempo. Paton and Buonomano (2018) gave the telling example of perception of messages in Morse code, where the meaning is determined entirely by differences in duration of successive stimuli, and individual dots and dashes acquire meaning from their temporal location in a hierarchical sequential structure. Without a timing mechanism that could represent both duration and temporal order, any information about a Morse code message in the EP would be unintelligible. That implies that components of information in the EP are marked with ordinal information and duration information.

Semantic encoding of temporal information is not a novel hypothesis. Dennett and Kinsbourne (1992) proposed that items of input information are marked with their time of occurrence. An example they used is somatosensory information: if a toe and a shoulder are stimulated simultaneously, input takes longer to reach cortical processing from the toe than from the shoulder (Efron, 1963; Vroomen & Keetels, 2010). Subjective simultaneity, they argued, is not a matter of information from the shoulder being held in a store until information from the toe arrives, but a matter of each piece of information being marked with its time of occurrence, which then forms part of the percept. As they said, this applies to temporal order percepts as well. Information that A occurred before B is not in the form of a percept of A followed by a percept of B (although that might well occur); it is in the form of a piece of information saying that A occurred before B. Time marking has also been proposed as an explanation for some phenomena in perceptual processing (Herzog et al., 2016; Nishida & Johnston, 2002).

To illustrate what is being proposed here, once the fourth chord of Beethoven's fifth symphony has been processed and has entered into the

EP, it is represented there with a number of features, among which is time marking information. The chord is marked as occurring immediately after the third chord (ordinal information) and as having a certain duration. This is not actual time: it is information about time. The information about the chord's duration does not itself have that duration, it just indicates what the duration was.

Given that the information structure in the EP may comprise multiple items with short temporal durations or ordinal relations, organised on the supra-second time scale, there is evidently a need for timing mechanisms for both duration and ordinality on multiple time scales. At least three different ways of registering temporal information may be relevant to the EP.

First, some temporal information is registered in perceptual processing, prior to entry of information into the EP. For example, Hogendoorn, Verstraten, and Johnston (2010) found evidence for multiple local (duration) timing mechanisms in the visual field, which could be desynchronised by local adaptation procedures. Mechanisms for assessing ordinal relations between events on short time scales clearly exist because many studies have shown thresholds for temporal order judgment as low as 20 ms (e.g. Brown & Sainsbury, 2000; Craig & Baihua, 1990; Eimer & Grubert, 2015; Fink, Churan, & Wittmann, 2005; Hirsh & Sherrick, 1961; Nicholls, 1994; Stevens & Weaver, 2005; Tadin, Lappin, Blake, & Glasser, 2010). Temporal order discrimination thresholds can range up to 100 ms and even longer (Craig & Baihua, 1990; Fink, Ulbrich, Churan, & Wittmann, 2006; Matthews & Welch, 2015; Nishikawa, Shimo, Wada, Hattori, & Kitazawa, 2015). So information enters the EP from perceptual processing already with local time marking, including ordinal information, on a scale possibly up to about 100 ms.

There is also evidence for timing mechanisms operating on the subsecond time scale. A particularly relevant example is state-dependent networks (SDN; Goel & Buonoamno, 2014; Wittmann, 2013). In a SDN, when a stimulus arrives there is an initial brief spike response from one or more neurons, followed by a more gradual change in inhibitory post-synaptic potential (IPSP). The state of the IPSP modifies the response of neurons in the network to the next stimulus, with some cells inhibited and some cells facilitated, and the degree of modification depends on the amount of time that has elapsed since the first spike response occurred. Because of that, the response to the second stimulus potentially yields information about the amount of time that has passed since the first stimulus (Buonomano & Merzenich, 1995; Karmarkar & Buonomano, 2007). The relevance of this to the EP is that, whereas most timing mechanisms that have been proposed are concerned with duration or interval timing, SDNs can function as registers of temporal ordinality: “if stimulus “A“ is presented to an animal, “A“ will produce a change in cortical network states as a result of time-dependent neuronal properties and stimulus “B“ will then produce a pattern of activity that codes for “B“ preceded by “A,“ rather than simple “B”“(Buonomano & Merzenich, 1995, p. 1030). Karmarkar and Buonomano (2007) proposed an output layer of neurons that effectively transform a temporal pattern of activity into contemporaneous information about the sequence of events. That information is then available to further processing. In short, the ordinal temporal information can be attached to items of information, and this could then be entered into the EP. The time scale of the model developed by Buonomano and Merzenich (1995) was about 300 ms, but the model can be extended to encompass shorter and longer time scales (Goudar & Buonomano, 2014). SDNs appear to model time marking on a time scale longer than that of perceptual processing and would therefore be suitable mechanisms for registering time information within the time scale of the EP.

There is also a need for time marking up to the maximum temporal extent of the EP. For example, a timing mechanism is needed to mark the differentiation between information structures in the EP, associating timing information with information structure boundaries (see next section for discussion of event boundaries in relation to the EP). There is evidence for ordinal and interval timing mechanisms on the supra-second scale (Cona & Semenza, 2017; Coull, Charras, Donadieu, Droit-Volet, & Vidal, 2015; Guidali, Pisoni, Bolognini, & Papagno, 2019; Kalm & Norris, 2017; Protopapa et al., 2019). But the specific requirement for the EP would be not an optional process for generating timing information on demand but an automatic, continuously operating mechanism involved in assembling the information in the EP. That is, however timing information may be generated, it becomes a semantic feature of a perceptual/memorial object just as other semantic information such as category membership become features of perceptual objects.

One possible mechanism is based on changes in neural firing rates over the time course of an event. It has been proposed that information about temporal interval and order is automatically captured by time-dependent changes in the state of neural networks (Mauk & Buonomano, 2004; Wittmann, 2013). Merchant, Harrington, and Meck (2013) reviewed evidence from several studies that neural circuits in various areas of the brain exhibit ramping activity during timing tasks, such as timing of motor responses. The ramping activity can be read off as an indicator of the passage of time. Noise in the response of single cells can be minimised by reading off a kind of running average over a population of neurons. In two studies there is evidence for a mechanism of this kind operating on a time scale commensurate with that of the EP. Wittmann (2013) and Wittmann et al. (2011) found evidence for ramping activity during temporal encoding of durations of 9 s and 18 s, localized in the posterior insula. Another study found evidence for ramping activity involved in the encoding of both duration and relative duration of two stimuli on time scales up to ~3 s (Genovesio, Tsujimoto, & Wise, 2009). A study looking at rather longer time intervals identified numerous cells that exhibited ramping activity at different rates in association with event boundaries; that is, the cells reset when a particular environmental event occurred (Tsao et al., 2018). This research shows timing mechanisms that identify event boundaries, so they appear to be timing the durations of events on the supra-second scale. All of these studies are consistent with the hypothesis that ramping activity in populations of neurons could provide timing information for the EP: together, they show duration and relative duration timing for multiple time scales on the supra-second scale, and also timing linked to event boundaries.

In summary, there is evidence for temporal order marking in perceptual processing, for registration of both duration and ordinal temporal information on the sub-second time scale with SDNs, and time marking in relation to the larger scale structure of the EP, including event boundaries, with ramping activity in populations of neurons. Between them, these three levels of time processing can provide all the time marking information that is needed to integrate information in the EP in a way that indicates temporal duration and ordinality on multiple time scales, and marks boundaries between one information structure and the next. Much more research is necessary to investigate the operation of these mechanisms in relation to information in the EP. It is one thing to look at timing mechanisms per se, as it were detached from the rest of processing in the brain, quite another to look at the assignation of time markers to sets of information currently being processed and held in the EP. This would be a key priority for future research.

4. The temporal boundary of the extended present

Just as a perceptual object is a collection of features integrated into a coherent, bounded representation, so the EP is a collection of memorial information integrated into a bounded information structure. On that hypothesis, what sets the temporal boundary of the EP?

Given that the key proposed features of the EP are thematic connection and temporal information, one obvious answer would be that the boundary of the EP is set by the termination of a thematic connection, which would often mean the termination of a hierarchical structure. In the case of language, for example, the boundaries would be set by the termination of a sentence or a complete grammatical unit, or perhaps by the initiation of the next sentence. Thus, the end of a sentence marks the temporally proximal boundary and retained information about the beginning of that sentence marks the temporally distal boundary. While this may be sufficient for locally complete hierarchical structures, many hierarchical structures have no fixed termination, or may continue for a long span of time and encompass a large amount of information. In an extended piece of music, for example, several minutes of music may all be thematically connected, and may all be analysable in terms of a single overarching hierarchical structure. The entire first movement of Beethoven's fifth symphony would be an example. In the case of making a cup of coffee, or watching another person doing that, multiple actions over a span of two or three minutes might all form part of a connected and hierarchically structured action sequence. If the EP can be demarcated as an information structure, it should be possible to identify local completions in temporal hierarchical structures that function as temporal boundaries to the EP.

One possibility is that the boundary is set by information storage capacity limitations. In the section on constructing hierarchical structure information in the extended present it was argued that integration cost is minimal for the EP because integration is accomplished by rapid automatic processes of high capacity, based on spreading activation. Memory storage cost (Gibson, 1998) is more likely to be a relevant factor. Individual WM stores have limited capacity: the capacity of the phonological loop, for example, has been estimated at two or three items (Baddeley & Hitch, 1974). However, effective storage capacity can be supplemented by involvement of LTM. Individual items can be bound into meaningful chunks with the aid of activated information from LTM (Norris, 2017; Wingfield & Butterworth, 1984). Information may also be transferred from WM to LTM where it is transiently high in accessibility due to strong associative bonds with other information retained in WM. The EP is a mutable, dynamic construction with information constantly being transferred from WM to LTM and activated from LTM to WM in accordance with ongoing processing tasks. Over a short period of time, activation of accessible information in LTM may increase the effective storage capacity of the EP well beyond the limits on individual WM stores. As other authors have noted, there may be no definable limit on storage capacity in WM (Caplan & Waters, 2013; Norris, 2017) and, for the same reasons, no definable limit on the storage capacity of the EP. Because of that, although capacity limits may determine boundaries to the EP to some extent, it may not be possible to define a limit in terms of storage capacity alone.

Another possibility is that, at least in the case of action, information pertaining to a sub-goal becomes redundant and need not be retained when the sub-goal is accomplished. This implies that there would be preferential retention of information about goals and sub-goals not yet accomplished. Thus, in the coffee-making example, once the sub-goal of filling the kettle with water is accomplished, the hierarchical structure of actions specific to filling the kettle is redundant and need not be retained in memory. This principle can be generalised to other domains. For example, local temporal hierarchical structure information about music need not be retained when a musical phrase boundary is identified, even though some information about the history of the music might still be required to make sense of what is going on.

Many studies have shown that, when observing an action sequence, people spontaneously segment the sequence into discrete events; there is reasonable consensus on where, in the stream of action, event boundaries are located; and segmentation can occur on many time scales (Kurby & Zacks, 2008; Radvansky & Zacks, 2017; Richmond & Zacks, 2017; Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zacks & Tversky, 2001). Most studies of event segmentation have used stimulus presentations with a quantity of information well beyond WM capacity, often on a time scale of minutes (Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015; Hard, Recchia, & Tversky, 2011; Zacks, Tversky, & Iyer, 2001). However, that does make the research suitable for identifying factors involved in setting boundaries in the EP because not all the presented information could be retained in the EP.

There is evidence that goal completion is a determinant of event segmentation (Baldwin & Baird, 2001; Kopatich, Feller, Kurby, & Magliano, 2019; Levine, Buchsbaum, Hirsh-Pasek, & Golinkoff, 2018; Zacks, 2004). For example, when stimulus information indicates a change in goals of one or more characters in a story, participants tend to identify an event boundary at that point (Magliano & Zacks, 2011). Levine et al. (2017) found evidence that event segmentation was associated with both goal completion and goal initiation. Infants aged six months have some understanding of goals in the perceived actions of other humans (Daum et al., 2009; Falck-Ytter et al., 2006; Luo, 2010). At the same age, infants are sensitive to violations of syntax in stimulus presentations of hierarchically structured goal-directed actions (Maffongelli et al., 2018). Before the age of one year, infants segment goal-directed action sequences in ways that align with the goals of the actors (Baldwin, Baird, Saylor, & Clark, 2001; Saylor, Baldwin, Baird, & LaBounty, 2007). By interweaving actions related to two different goals, Loucks, Mutschler, and Meltzoff (2017) showed that children aged three years represented the events in terms of the goal structure, even if that meant violating the temporal order of the action sequence. This indicates an understanding of events in terms of goal structures, and events are segmented accordingly.

If the hierarchical structure account of boundary identification in the EP is correct, it implies that information on the far side of a structure boundary, or information within a larger goal structure that has become redundant, is actively removed or erased from the EP. There is indeed evidence for this. Removal of information was identified by Lewis-Peacock et al. (2018) as a control process involved in the operation of WM, erasing outdated information from WM to reduce WM load and prevent or minimize interference with the current task. They reviewed several lines of evidence for removal: “(1) an improvement of performance in WM tasks after irrelevant information has been removed; (2) reduced access to the removed information; and (3) reduced neural activity correlated with removed information” (p. 34). Several other studies have found support for removal (Oberauer, 2001, 2018; Souza, Rerko, & Oberauer, 2014; Williams, Hong, Kang, Carlisle, & Woodman, 2013). In the event segmentation literature, Speer and Zacks (2005) found that, when an event boundary is identified in narratives, memory for information beyond the event boundary suffers a decrement, consistent with the hypothesis of an active removal process.

In application to the EP, removal of redundant information on the far side of an information structure boundary is functional because it releases storage space and processing resources for construction of a new set of information in the EP. The same applies to removal of information about redundant sub-goals such as the filling the kettle sub-goal once the kettle has been filled. Removal need not mean total erasure, but it does mean that information about or in the previous EP must be stored in LTM if it is to survive in the system. It may still be transiently high in accessibility from LTM depending on the thematic connection (associative strength) with the information in the new EP, but it is likely to be selectively reduced and transformed. As an example, the research on music processing shows that, beyond the reach of the EP, hierarchical structure information is largely lost but some general information about the earlier parts of the piece of music, such as global patterns of tension development and resolution, may be retained and accessible from LTM (Deliege, 2001; Granot & Jacoby, 2011; Lalitte, Bigand, Kantor-Martynuska, & Delbe, 2009; Tillmann & Bigand, 2004).[4]

In the event segmentation literature, there is evidence for other cues to event boundaries. These include relatively large amounts of change (Michaels & Carello, 1981; Newtson, Engqvist, & Bois, 1977; Rosch,

---

[4] The removal process discussed here also resembles the “chunk-and-pass” processing mechanism proposed by Christiansen and Chater (2016) as a means of coping with the processing capacity demands associated with linguistic input and the consequent rapid loss of information. However, the processing bottlenecks with which they were concerned occur prior to entry of information into WM, in perceptual processing and sensory memory.

1978) lack of or reduction in amount of change (Hilton, Räling, Wartenburger, & Elsner, 2019), and prosodic cues such as slowing of tempo and pitch rise (Hilton et al., 2019). In fact all of these are associated with completion of temporal hierarchical structures. In action sequences, amount of change tends to be correlated with goal or sub-goal completion, and it may be perceived goal completion and not amount of change that determines the setting of the event boundary (Baldwin & Baird, 2001; Levine et al., 2018). Take, as a simple example, an action sequence involving reaching for, grasping, and lifting a mug of coffee. This can be treated, for illustrative purposes, as involving two sub-goals, reaching for the mug and lifting it. The reaching phase involves continuous motion of the arm and hand with minor and smooth changes to kinematic features. When the sub-goal of reaching the mug is accomplished, there is much change: adjustment of fingers into a gripping position, initiation of motion of the mug as it is lifted, and a new motion trajectory being initiated for arm and hand. After that, only minor changes in trajectory occur again. In that example, amount of change is closely associated with sub-goal accomplishment. Reduction of change in the form of pauses or rallentando is associated with hierarchical structure boundaries in music (Cuddy et al., 1981; Palmer & Hutchins, 2006) and in language (Frazier et al., 2006). And it was shown in the section on hierarchical structure processing that prosodic hierarchies tend to be closely associated with hierarchical structure boundaries (Heffner & Slevc, 2015). Therefore, while amount of change and prosodic features might be determinants of event boundaries, it is also possible that they are merely associated with hierarchical structure boundaries, and hierarchical structure boundaries determine the boundaries in the EP.

The main hypothesised alternative to structure completion as a determinant of event segmentation is prediction error. Zacks et al. (2007) proposed a model of event segmentation that combines perceptual processing with predictions based on event models, and a prediction error detection process. The model generates predictions for immediate future events; these are held in a comparator which compares them against sensory input about the actual events; if a discrepancy exceeding some criterion value is detected, an error is signalled. This results in a change to the predictive model. Prediction in general, and comparator models in particular, are important components of models in many domains of processing (Bar et al., 2006; Blakemore, 2003; Blakemore, Smith, Steel, Johnstone, & Frith, 2000; Clark, 2013; Hohwy, 2013; Khoei, Masson, & Perrinet, 2017; Pickering & Gambi, 2018; Pickering & Garrod, 2007; Richmond & Zacks, 2017; Wolpert, Ghahramani, & Jordan, 1995). The question for present purposes is not whether prediction error detection occurs but whether it is used as a cue to event segmentation.

The hypothesised role of prediction error in event segmentation has been discussed extensively by its authors (Radvansky & Zacks, 2017; Richmond & Zacks, 2017; Zacks et al., 2007), but there is not much supporting evidence. The results of a study by Avrahami and Kareev (1994) were cited as support for the prediction error hypothesis (Radvansky & Zacks, 2017). Avrahami and Kareev (1994) found that subsections of a video sequence that had been perceived several times previously with different sub-sections preceding and following them were identified as distinct events. Familiarity seems to be the sole basis for this segmentation. Possibly the end of a familar sub-section marks a moment of prediction error. That, however, would be specific to stimuli where familiar sub-sections were embedded in unfamiliar ones, and it is not clear that the finding would generalise beyond that case.

If goal completion can be predicted, and yet is still used to identify an event boundary, then prediction error cannot be involved in that. That would seem to be the case for many familiar action sequences, such as making coffee. Related to that, Clewett and Davachi (2017) pointed out that the prediction error hypothesis breaks down when a novel event sequence is encountered: event segmentation still occurs, even though perceivers have no basis for making predictions. The results of a study of event segmentation of narratives by Magliano, Kopp, McNerney, Radvansky, and Zacks (2012) were also cited as support for the prediction error hypothesis (Radvansky & Zacks, 2017). However the results showed that event segmentation was based on change in significant situational features, which were defined as including entities (e.g. people), goals, causal relations, and spatial and temporal information. There is no indication in the study as to whether these changes involved prediction errors or not.

Zacks, Kurby, Eisenberg, and Haroutunian (2011) presented films of action sequences with durations of a few minutes. At various points the film was stopped and participants were asked to predict what would happen in the next few seconds. The stopping points were either just prior to an event boundary or in the middle of a segment, as identified by participants in a previous experiment. Predictions were less accurate at event boundaries than in the middle of segments. This is consistent with the prediction error hypothesis. However, it does not establish that prediction error was the cue for segmentation. Suppose that goal completion is the cue for segmentation. Then it follows that prediction will be relatively easy in the middle of a segment because progress towards the goal is continuing (as in the example of reaching for a coffee mug), but will be relatively difficult at the completion of a goal because there is hardly a limit on the number of possible new goals that might be set up. Thus, prediction error is associated with goal completion, so the study does not establish that prediction error and not goal completion was the effective cue for event segmentation. Zacks et al. (2011) ruled out amount of change in action and amount of change in image similarity as alternative possible interpretations, but they did not test the goal completion hypothesis.

In summary, the evidence from event segmentation research favours goal structure as a key determinant of event segmentation. Since goal structure is a kind of hierarchical structure, this is consistent with the proposition that the boundary of the EP is set by hierarchical structure boundaries. Other cues such as relative amount of change and prosodic features may function as cues to hierarchical structure boundaries and subserve the function of boundary setting in that way. Given that hierarchical structures may be very substantial (e.g. symphonic movements), it is likely that goal structure is moderated by storage capacity limitations. It is also possible that redundant hierarchical structure information, such as the structure of a sub-goal that has been completed, is actively removed (Lewis-Peacock et al., 2018).

5. Functional significance of the extended present

The present paper has been concerned with defining the EP and this has entailed detailed scrutiny of research literatures relevant to, and providing evidence for, its proposed features. Brief observations about its functional significance are appropriate, however, because it could turn out to be of considerable importance in cognition.

The starting point for this paper is the idea that the EP helps to make sense of what is going on now. Suppose that there is no EP. Then what sense do we make of the eighth chord of Beethoven's fifth symphony, or of the sixth word in a sentence when we hear it? The answer would be that the chord or word would have its own individual meaning, and it might even be perceived as part of a sequence of events to the extent that they are retained in WM, but that would be all. The overall meaning of the sentence, and the specific meaning of the word in that particular sentence, would not be there. The role of the eighth chord in the phrase structure of the symphony up to that point would not be represented, just its position in a temporal order of individual events. We would not experience music as music any more. Without temporal order information and hierarchical structure information the sequences would not even be perceived as sequences, but just as isolated events with no temporal relation to each other. Retaining information about the order of events in time is possibly the most neglected function of short-lasting memory, and it is fundamental to the EP.

But there may be more to the function of the EP than that. The EP could be an enabler for both execution and perception of complex operations in language, action, music, exact calculation, and possibly other domains of cognition. That is, it is an information structure with particular properties, and the information in it can be accessed and utilised by specialised processors. The mere capacity to hold a certain quantity of information in an activated state in WM is not sufficient to render linguistic utterances and goal-directed actions comprehensible. The particular features of the EP are necessary for that. The EP is a specialised representation of thematically and temporally connected information. As such, it is an enabler for cognitive operations that construct and operate on thematically and temporally connected information. Having a store with just those specialised features makes it possible to comprehend and to generate grammatically complex utterances, extended sequences of goal-directed actions, music and, possibly, complex arithmetical operations. The thematic connection with present perceptual input is necessary for the construction side of those cognitive operations, in the sense that present perceptual input may lead to modification of the entire developing construct. Without a store that had the particular capabilities of the EP, that would not be possible.

In general, memory stores have the function of holding information in a form that is accessible to further processing. Any process that can take the kind of information that is in the EP can benefit from it. As a hypothetical example, that might include metacognition. For example, it would be possible to hold a temporally and hierarchically structured representation of part or all of an action sequence in the EP, perhaps aided by rehearsal and transiently accessible storage in LTM, and operate on it with metacognitive processing. That could support innovation in the form of change to some part of the action sequence, which could then be tested and further refined. An example of such a change would be substitution of a new means of achieving a sub-goal that was more efficient than the existing one but still fitted with the overall action sequence. That kind of metacognitive operation would not be possible without the particular structure of information that is held in the EP.

6. Comparison with “subjective present” proposals

Other authors have claimed that there is a “subjective” or “specious” present on a time scale of a few seconds (Fraisse, 1984; James, 1890; Michon, 1978; Montemayor & Wittmann, 2014; Pöppel, 1997, 2009; Wittmann, 1999, 2009, 2011, 2013; Wittmann & Pöppel, 1999/2000). Proposed durations for the subjective present range from 2 to 3 s (Pöppel, 1997, 2009) to a maximum of 10 s (Clarke, 1987). Generally, it is proposed that there is global temporal integration of information on that kind of time scale, as if the whole of cognition was divided into units of a few seconds, within which all information is integrated. Pöppel (2009) wrote, “normally after an exhaust period of 2 - 3 s, attentional mechanisms are elicited that open the sensory channels for new information. Metaphorically speaking, every 2 - 3 s, the endogenously generated question arises ‘what is new’“ (p. 1893). And: “the brain creates temporal windows of just a few seconds within which the identity of a percept or a concept is maintained (stationarity), and allows after such an interval the access of a new percept or concept” (p. 1894). The processes that generate the subjective present are supposed to be automatic (Pöppel, 1997, 2009). Thus, the evidence that would support the proposal of a subjective present would show automatic temporal integration in multiple domains all operating on approximately the same time scale. That argument was made by Pöppel (1997, 2009) and Wittmann (Montemayor & Wittmann, 2014; Wittmann, 1999, 2009, 2011), and they surveyed research on several kinds of automatic processes, claiming that they did indeed share a common time scale of temporal integration.

The topics cited by those authors as relevant to the proposed subjective present included accuracy in reproduction of stimulus durations, synchronization of behaviour with a regular beat, mental rhythmization of a regular beat, time units in behaviour, time scale of reversals of perception with bistable ambiguous figures, time scale of inhibition of return in visual search, and EEG responses to deviant stimuli in series of repeating stimuli. Research on those topics was reviewed by White (2017) and was found not to support the proposed global integration, as was discussed in the introduction above.

What the present proposal has in common with the previous ones is the idea that there is an integrated body of recent historical information. In the present proposal that integration is accomplished by mechanisms of hierarchical structure processing, temporal processing, and identification of boundaries, with removal of information on the far historical side of a boundary. It is not the case that all information active in cognition over a short time span is integrated; on the contrary, integration is confined to thematically connected information and does not extend to other information that is in WM but not thematically connected to what is in the EP. The previous proposals sought evidence for temporal integration within domains: the argument for a domain-general representation depended on the claim of a common time scale of integration across the domains in question. The review by White (2017) showed that the common time scale of integration was lacking, and there was no evidence for overall or domain-general integration.

An example would be bistable ambiguous figures, where research shows spontaneous shifts between one possible perceptual interpretation of the figure and another (e.g. Orbach, Ehrlich, & Heath, 1963). Pöppel (1997, 2009) and Wittmann et al. (2011) claimed that shifts from one percept to the other occur on a time scale of approximately 3 s, and claimed in addition that this was supportive evidence for the proposal of a subjective present. However, theoretical models of perceptual alternations have focussed on domain-specific phenomena such as fatigue or adaptation (Furstenau, 2014; Lee, 2014) that would have no application to most of the evidence cited in support of the subjective present proposals (White, 2017). Indeed, perceptual alternation rates for bistable stimuli in different modalities are not correlated across participants (Wernery et al., 2015), which disconfirms the hypothesis of a common temporal integration mechanism even within the domain of bistable perceptual alternations. Moreover, if Pöppel's (2009) metaphorical ‘what is new?’ question is asked, the answer in the case of a figure with bistable perceptions is “nothing” because the figure (e.g. a Necker cube) has not changed.

In summary, the research evidence cited as supporting the hypothesis of a subjective or specious present actually shows either different, domain-specific forms of temporal integration or other phenomena, such as adaptation. Information integrated in such ways could be input to the EP but does not contribute to an explanation of how the EP is constructed.

7. Time scale of the EP

Previous proposals have defined the subjective present, at least in part, in terms of the time scale of information encompassed by it, with time scales ranging from 2 to 3 s (Pöppel, 1997, 2009) to 10 s (Clarke, 1987). The present account has avoided the issue of the time scale of the EP, arguing that the boundary to the EP is set by processing considerations such as hierarchical structure completion and storage capacity limits, not by a unit of time. Those considerations do permit a little to be said about time scale, though it is necessary to repeat, first, that time scale does not define the EP.

The most recent end of the EP would be a single item of information temporally proximal to the current perceived present. Thus, if one is currently hearing the second chord of Beethoven's fifth symphony, meaning that that chord is in the EP, then the retained information about the first chord is the full temporal extent of the EP at that moment. This implies that the shortest time scale would be the minimum temporal order discrimination threshold. Research evidence surveyed in the section on termporal information above indicates that this is no more than 20 ms. If two auditory clicks are presented 20 ms apart, when the second one is in the perceived present, the first would be held in sensory memory, along with time marking information that relates it to the second click. Of course with two clicks there is no hierarchical structure, just a temporal sequence, so perhaps 40 ms would mark the minimum time scale on which hierarchical structure information could be represented.

The distal end of the time scale is impossible to define in terms of time. It is defined by completion of a hierarchical structure thematically connected to the perceived present and by information storage capacity limits, which are themselves impossible to define because of the possibility of transiently accessible storage of information in LTM. To continue with the music example, probably the simplest musical structures are exemplified by some forms of chant consisting of monodic strings of long-duration notes. If 16 chords of a Bach chorale prelude can be held in the EP (Koelsch et al., 2013) then it would certainly be possible for 16 notes of a monodic chant to be held there, and the time span of that could be 20 s or more. Perhaps under conditions of excellent concentration, minimal distraction, and simple hierarchical structures with few elements each of long duration, the distal end of the EP could be more than 20 s into the past. If so, then the EP could have a time scale ranging from 40 ms to 20 s or more. Of course, the extremes of 40 ms and 20 s would rarely be met with in practice. The typical duration of the EP would probably be more in the range of about 300 ms (typical syllable length) to a few seconds, probably not dissimilar to the time scales proposed for the subjective present. What matters, however, is that, whatever the typical time scale of the EP might be, the EP itself is not defined in terms of time, and the kind of integration of information proposed here (hierarchical structure plus temporal information) is different from the kinds of temporal integration that have been proposed by other authors (Fraisse, 1984; James, 1890; Michon, 1978; Montemayor & Wittmann, 2014; Pöppel, 1997, 2009; Wittmann, 1999, 2009, 2011, 2013; Wittmann & Pöppel, 1999/2000).

8. Is there a basic unit of information in the EP?

In the proposal by Pöppel (1997, 2009) the three-second moment proposal was accompanied by a smaller level of temporal integration on a time scale of about 30 ms. Previous reviews have not supported that proposal, showing great variation in time scales of temporal integration, and temporal units that are domain-specific (White, 2017, 2018). However, if hierarchical structure is a key feature of the EP, perhaps it could still be argued that the lowest level of the hierarchical structure comprises base units that share a short, common time scale.

Rajendran, Teki, and Schnupp (2018) have argued that domain-general timing mechanisms set values of fundamental temporal parameters in the perception and production of rhythmic patterns. Starting with an analysis of the time scale of rhythms in music, they argued that “the timescales that are relevant in music are also relevant in other contexts such as in the production and perception of movement (walking, running, breathing) and speech, and in the parsing of complex acoustic scenes” (pp. 13–14). The time scale in question is about 400–600 ms. Rajendran et al. stated, “When asked to judge the duration of time intervals, there is a systematic tendency for human listeners to overestimate shorter time intervals (roughly 250-400 ms) and underestimate long ones (~600 ms to 2 s)” (p. 5). The interval of accurate judgment in between is referred to as the indifference interval, and they argued that it corresponds to a comfortable walking pace and also the rate at which people spontaneously tap. If that is right, then units of about 400–600 ms may be the basis for construction of larger-scale information structures and behavioural patterns. The initial evidence is not encouraging: reviewing evidence published up to that date, Van Noorden and Moelants (1999) argued that the value of the indifference interval depends on the method used to study it and that it tends to be around the mean of the durations used in any given experiment, with a large range of values (up to 3000 ms reported by Pöppel, 1978). Also, although Rajendran et al. (2018) included breathing in their argument, the average length of a single normal breathing cycle (inspiration + expiration) when resting is about 4 s (White, 2017), which is far beyond the indifference interval. The possibility of a domain-general temporal base unit is worthy of investigation, however.

Research on rates of several kinds of behaviour was surveyed. The results are shown in Table 1. Table 1 does show some rates falling into the range 400–600 ms, including a study of grouping in subjective rhythmization (Parncutt, 1994), mean beat frequency in recordings of popular dance music (Van Noorden & Moelants, 1999), spontaneous tapping (Repp, 2005), step rate in walking (Macdougall & Moore, 2005; Sardroodian et al., 2015), and bouncing while lying on an inclined sled (Raburn et al., 2011). The table also shows that most rates fall outside that range, so overall the results do not support the hypothesis of a base unit of time scale in behaviour. The research summarised in Table 1 was focussed on the shortest time scales of behavioural units. Other repetitive behaviours, such as chewing and breathing, have mean repetition rates well above 1000 ms (Gerstner & Cianfarani, 1998; Po et al., 2011; White, 2017), further undermining the case for a basic temporal unit in

Table 1
Base units for various kinds of behaviour.

Mean	Paper
Grouping in subjective rhythmization	
1100 ms	van Noorden and Moelants (1999)[a]
558 ms	Parncutt (1994)
~1500 ms	Szelag, von Steinbüchel, Reiser, de Langen, and Pöppel (1996)[b]
1580 ms	Bolton (1894)[c]
~1500 ms	Baath (2015)[c]
Mean beat frequency in recorded dance music	
447–499 ms	van Noorden and Moelants (1999)[d]
Spontaneous tapping rate	
500–600 ms	Repp (2005)
240 ms	Hansen and Ohnstad (2008)
331–368 ms	Sardroodian, Madeleine, Mora-Jensen, and Hanse (2016)[e]
418 ms	Wittmann et al. (2001)
Step rate in walking	
500 ms	MacDougall and Moore (2005)
520 ms	Sardroodian, Madeleine, Voigt, and Hansen (2015)
Step rate in running	
370 ms	Sardroodian et al. (2015)
Cycle rate in bicycle riding	
350 ms	Hansen and Ohnstad (2008)
380 ms	Sardroodian, Madeleine, Voigt, and Hansen (2014)
390 ms	Sardroodian et al. (2015)
Bouncing (lying on a sled at a 60° incline)	
526 ms	Raburn, Merritt, and Dean (2011)
Sawing cycle (complete forth and back movement) rate	
333–1000 ms	Starke and Baber (2017b)[f]
Arm stroke rate in swimming	
1282 ms	Chatard, Collomp, Maglischo, and Maglischo (1990)[g]
1041–1413 ms	Pelayo, Sidney, Kherif, Chollet, and Tourny (1996)[h]
1453–1554 ms	McCabe and Sanders (2012)[g]
Speech rate	
128–193 ms	Pellegrino et al. (2011)[i]
151 ± 31 ms	Coupé et al. (2019)[j]
320 ms	Vollrath, Kazenwadel, and Krüger (1992)[k]

[a] For stimuli with ISI ~600 ms groups of two were likely to be perceived; for ISI ~300, groups of four, and for ISI ~ 150 ms, groups of eight.

[b] Incorporating elements into a single unit of rhythm, upper limit.

[c] Upper limit of intervals between beats for subjective rhythmization.

[d] Mean beat frequency in popular dance music, six countries sampled.

[e] Mean tapping frequency increased over two weeks of testing.

[f] Length of blade accounted for the range found; other variables had little or no effect.

[g] Freestyle, competitive swimmers.

[h] Freestyle, competitive swimmers at distances from 50 m to 1500 m. Stroke rate decreased as distance increased.

[i] For syllables, range for seven languages.

[j] For syllables, mean and SD for seventeen languages in nine families.

[k] Phonation (vowel-consonant unit) rate.

perception or behaviour.

The main problem is that time scales in behaviour are determined by biomechanical and bioenergetic considerations more than by brain processes. Starke, Baber (2017b) cited research showing that "repetitive actions are generally performed at rates which minimize energy consumption" (p. 2). Minimal energy consumption rate varies depending on several factors. An example would be the length of the effector units, so that finger tapping tends to occur at a faster rate than walking as a biomechanical consequence of the fact that legs are longer than fingers. External circumstances also affect rate. Thus, arm stroke rates in swimming tend to be slower than step rates in walking and running, and cycle rates in bicycle riding (see Table 1), because of bioenergetic considerations involved in pushing the arms through water, which offers more resistance than air. Sawing rates depend simply on the length of the blade, and therefore the permissible range of movement (Starke, Baber, 2017b). Different rates of behaviour can even occur simultaneously: Starke and Baber (2017a) found that people can saw at their natural rate with one hand and tap at a different rate to a metronome with the other.

The relevance of biomechanical and bioenergetic considerations in behaviour might suggest that perception is a better place to look for base units of temporal integration. That literature was reviewed by White (2018). Evidence of a great range of temporal integration times was found, ruling out the hypothesis of a domain-general unit of temporal integration. There may well be domain-specific consistencies in temporal integration units, and it is likely that at least some of that is driven by periodic oscillations in brain activity revealed in EEG recordings (e.g. VanRullen, 2016; VanRullen & Koch, 2003; VanRullen, Zoefel, & Ilhan, 2014). External factors are still relevant in perception, however. Processing of speech input is constrained by natural variation in the duration of units at all levels of analysis. Syllables, for example, vary in duration from ~125 - ~400 ms (Hickok & Poeppel, 2007; Poeppel, 2003; Poeppel & Assaneo, 2020), and perception of these and other linguistic units must involve temporal integration over their naturally varying time scale, and not run on some fixed base unit. Speech may be (and be perceived as) quasi-rhythmic but is not periodic (Poeppel & Assaneo, 2020). And the time scale of units at the level of syllables is specific to speech, reflecting coordination in the movements of the vocal tract articulators (Poeppel & Assaneo, 2020) and not necessarily shared by units in other domains (White, 2017). If anything, given that repetition at a regular periodicity is more characteristic of behavioral output than of perceptual input, a domain-general base unit of temporal integration is less likely to be found in perception than in behaviour.

The EP, therefore, can accommodate structural information founded on variable base units of temporal integration; indeed, within reason, the time scale of the informational units of temporal integration is not relevant.

Having said that, it is possible to define basic units of information in the EP in another way. Take language as an example. In speech perception there are multiple time scales of analysis from basic distinctive features such as time between air release and onset of vocal cord vibration, through phonemes, syllables, words, and even beyond, with time scales ranging from 20 to 50 ms to many seconds (Hardy & Buonomano, 2016; Hasson, Chen, & Honey, 2015; Hickok & Poeppel, 2007; Poeppel, 2003; Rosen, 1992). Identifiable segments in lexical items, with information encoded in temporal windows of 20–50 ms (Hickok & Poeppel, 2007; Rosen, 1992) would seem to be the base level. However, perceptual processing of auditory input does not end at that level. Indeed a single segment on that time scale may not be sufficient to identify the sound as speech (as opposed to a wordless exclamation or the first note of a birdsong; Hickok & Poeppel, 2007). Perceptual processing integrates information over a longer time scale; in vision, temporal integration windows of 200–300 ms have been found (van Wassenhove, Grant, & Poeppel, 2007) and some visual perceptual products emerge from integration over more than 1000 ms (Burr & Santoro, 2001; Neri, Morrone, & Burr, 1998). Spoken syllables typically have a duration of approximately 150–400 ms (Hickok & Poeppel, 2007), most of which is within the temporal limits on integration. Hickok and Poeppel (2007) made a case for integration on both segmental and syllabic levels in perceptual processing, with parallel processing of both levels in the construction of perceived words.

In general, the products of perceptual processing are basically individuated, identified, bounded perceptual objects (not forgetting additional information about spatial relations and context), and those are entered into WM. Thus, in the case of speech perception, the basic unit of information in the EP would be words, constructed in perceptual processing by segmental and syllabic analysis on a time scale up to hundreds of milliseconds, if not more. This is not a rule without exceptions: for example, unfamiliar polysyllabic words such as "antidisestablishmentarianism" may pose problems for speech perception and may be entered into WM in syllables or short groups of syllables. Mostly, however, words are the units and syntactic structures in the EP are constructed from those units. In music the perceptual objects would be individual notes, chords, and rests, although long duration of some notes and chords might exceed the limits on temporal integration in temporal processing and be entered into the EP one segment at a time. In action it is reasonable to speculate that the base unit corresponds to the level of words in language, which would be specifically verbs. Thus, a grasping action comprises many individual motor commands and many individual muscle movements, but these are integrated into a bounded, unitary percept that is captured in language by the single verb "grasp" (Gibson, 1966).

To summarise, the basic level of units in the EP is individuated, identified, bounded perceptual objects, the constructs of perceptual processing that are entered into WM. These are most often, but not invariably, words, notes/chords/rests, and actions that can be described by single verbs.

9. Conclusion

Perceptual processing constructs information about what is going on in the present, with due allowance for processing latencies. In this paper it has been argued that present perception occurs in an informational context that plays a vital role in making sense of it. What is proposed is a synthesis of separate components into a coherent structure of information held in an activated or easily accessible state in WM and LTM, respectively. The information structure is held together by binding functions subserved primarily by thematic connection, temporal information, and hierarchical structure information. Taking a linguistic example, in the sentence, "the boy closed the door", "closed" is understood not just as a verb describing a particular action but as an integrated component of a sentence that has a coherent overall meaning (i.e. it is thematically connected to the rest of the sentence), a temporal location in the information structure of the sentence (i.e. it is the third word, occurring after "the boy" and before "the door"), and as occupying a place in a hierarchical structure. Those can all be understood as semantic labels accompanying the representation of "closed" in the EP and as tying it in particular ways to the other words in the sentence. New perceptual input that is thematically connected to the information structure of the sentence in the EP, such as the word "quietly", is interpreted in terms of its temporal and hierarchical relationship with the existing information structure.

A large part of this paper reviewed evidence relevant to the proposal. Most of the research reviewed concerns hierarchical structure processing, not because that is any more important than the other main features of the EP but because there has been more relevant research on that feature than on the others. Most of the evidence is consistent with what is proposed here, though it is evident that much remains to be done. Priorities for future research on the EP would primarily concern timing mechanisms. The importance of timing information, especially ordinal timing information, to information in the EP (and WM in general) has not been fully appreciated, and it seems to be taken for granted in much research to which it is actually profoundly important. It would also be important to develop a more nearly complete understanding of the nature of thematic connection and thematic boundaries between information structures: given that the flow of perceptual input is continuous, understanding how it is divided into meaningful temporal segments is important to the EP, but the time scale of the EP is shorter than that studied in most of the event segmentation research.

What novel predictions for research might follow from the proposed EP? The EP is a proposal about the ground that was covered by previous subjective present proposals. For example, paraphrasing Wittmann (2009), the subjective present is characterised as a perceptual unit with a duration of a few seconds and with temporal integration of the information it contains. Given that it is really a perceptual/memorial unit, since perception does not encompass that time scale, the question is what form of temporal integration is involved in the information that is in it? According to the papers on the subjective present, it includes such things as subjective rhythmization and its time scale, time units in speech and action, reversals of perception in bistable ambiguous figures, and the other topics included in my review paper (Pöppel, 1997, 2009; White, 2017; Wittmann, 2009). According to the EP proposal it includes hierarchical structure information and time marking information, with a distal boundary set by definable closure events such as accomplishment of a goal or sub-goal. These competing predictions can be tested by examining active information to see whether it is of the sort predicted by the EP proposal or the sort predicted by other subjective present proposals. The predictions made by the other subjective present proposals have been assessed and found to be wanting on the grounds that there is no common time scale of integration across the topics claimed to be relevant to the subjective present, indicating that their operation is determined by local processing considerations and not generalisable across the whole of the subjective present (White, 2017).

Viewed in that context, the basic prediction of the EP proposal is that the subjective present should contain hierarchical structure and time marking information that is thematically connected to current perceptual products, and its distal boundary is defined by closure events such as (perceived or performed) accomplishment of a goal or sub-goal. The problem with testing that hypothesis is that not everything that is in WM is also in the EP, so it is not enough just to examine the informational contents of currently active information. One way forward would be to present information of the sort hypothesised as being in the EP, as has been done in studies such as that by Koelsch et al. (2013). Given the importance of thematic connection to current perception, it would be predicted that current perception would be altered in predictable ways if the information presented in the recent past were interfered with - for example, if time marking malfunctioned or was made to do so, or if hierarchical structure analysis was interfered with (as was done by Maffongelli et al., 2015, for example). Equivalent disruption to things that are not in the EP should not affect current perceptual interpretation, or at least much less so. Take the example of a digit that was presented five seconds ago in a digit span test: that digit is stored in WM as an isolated item and is not in the EP. Current perception should not be significantly affected by knocking out or interfering with such isolated items, even large numbers of them.

Running a test of the EP hypothesis would not be straightforward because the EP cannot be reduced to its components. Showing that interfering in some way with hierarchical structure analysis affects current perception would not be enough because it would concern hierarchical structure analysis on its own. Possibly a combined assault could be useful. This might include: (i), presenting a sequence of stimuli with hierarchical structure, manipulating one component of the information, and assessing its effect via participants' responses to the terminal item (as in the study by Koelsch et al., 2013); (ii) combining that with manipulation of time order of components, to test hypotheses about the importance of time marking in representation of the information; (iii) use of brain imaging methods, since processing of the stimulus information should involve simultaneous activation of areas of the brain known to be involved in hierarchical structure analysis and temporal information processing. Cognitive load tasks would probably suffice to affect hierarchical structure analysis but time marking is a more fundamental, automatic process. In the present proposal, temporal information is set by basic features of neural activity, such as ramping activity in neural circuits, so it might seem difficult to manipulate the operation of such processes. However, misrepresentation of temporal order can be induced by experimental manipulations (Bechlivanidis & Lagnado, 2015; Cunningham, Billock, & Tsou, 2001; Stetson, Cui, Montague, & Eagleman, 2006), so it should be possible to take the next step and investigate how representations of objectively similar event sequences with different temporal order information in the EP would influence interpretation of current perceptual input. Research on how information in the EP is used in processes that have access to it would also be important. It should be emphasised, however, that the case for the EP would be made not by one or two results in specific areas of research but by the overall pattern in studies on all the hypothesised components of the EP. The present paper has attempted that as far as the current state of the research literature allows, but much remains to be done.

What matters most about the EP is that it is a coherent body of information, with all the parts of it connected. Temporal and structural information are the cement that binds the components together. Without that we would be lost in an endless, meaningless succession of isolated events. The extended present saves us from that and, to a considerable extent, makes experience meaningful. The “presentness” of the EP is not so much the point: what it provides is global integration of the products of local perception and integration processes that is responsible for the subjective coherence of experience and the general sense of ongoing happening.

Declaration of competing interest

I declare that there is no conflict of interest relating to this submission.

Acknowledgement

I am grateful to Candice Morey for helpful comments and advice on working memory.

References

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex, 21*, 1507–1518.

Amer, T., Gozli, D. G., & Pratt, J. (2018). Biasing spatial attention with semantic information: An event coding approach. *Psychological Research, 82*, 840–858.

Asano, R., & Boeckx, C. (2015). Syntax in language and music: What is the right level of comparison? *Frontiers in Psychology, 6*.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *Psychology of learning and motivation: Advances in research and theory* (pp. 89–195). New York, NY: Academic Press.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition, 53*, 239–261.

Baath, R. (2015). Subjective rhythmization: A replication and an assessment of two theoretical explanations. *Music Perception, 33*, 244–254.

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society B, 302*, 311–324.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Vol. 8. Recent advances in learning and motivation* (pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. (2019). The phonological loop as a buffer store: An update. *Cortex, 112*, 91–106.

Balaguer, J., Spiers, H., Hassabis, D., & Summerfield, C. (2016). Neural mechanisms of hierarchical planning in a virtual subway network. *Neuron, 90*, 893–903.

Baldwin, D., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708–717.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*, 171–178.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämälainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E.

(2006). Top-down facilitation of visual recognition. *Proc. Natl Acad. Sci. USA, 103*, 449–454.

Bechlivanidis, C., & Lagnado, D. A. (2015). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition, 146*, 58–66.

Bigand, E. (1997). Perceiving musical stability: The effect of tonal structure, rhythm, and musical expertise. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 808–822.

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience, 8, No. 94*.

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 184–197.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception and Psychophysics, 59*, 1098–1107.

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 159–171.

Blakemore, S.-J. (2003). Deluding the motor system. *Consciousness and Cognition, 12*, 647–655.

Blakemore, S.-J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine, 30*, 1131–1139.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150–171.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*, 145–238.

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences, 40*, 1+.

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language, 86*, 168–213.

Brown, L. N., & Sainsbury, R. S. (2000). Hemisperic equivalence and age-related differences in judgments of simultaneity to somatosensory stimuli. *Journal of Clinical and Experimental Neuropsychology, 22*, 587–598.

Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology, 76*, 30–77.

Buonomano, D. V., & Merzenich, M. M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. *Science, 267*, 1028–1030.

Burr, D. C., & Santoro, L. (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research, 41*, 1891–1899.

667 Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences, 21*.

Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory: Insights from neurodevelopmental and neurological disorders. *Annals of the New York Academy of Sciences, 1424*, 155–165.

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin and Review, 20*, 243–268.

Chait, M., Greenberg, S., Arai, T., Simon, J. Z., & Poeppel, D. (2015). Multi-time resolution analysis of speech: Evidence from psychophysics. *Frontiers in Neuroscience, 9*.

Chatard, J. C., Collomp, C., Maglischo, E., & Maglischo, C. (1990). Swimming skill and stroking characteristics of front crawl swimmers. *International Journal of Sports Medicine, 11*, 156–161.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language, 52*, 144–169.

Chen, L., Wu, J., Fu, Y., Kang, H., & Feng, L. (2019). Neural substrates of word category information as the basis of syntactic processing. *Human Brain Mapping, 40*, 451–464.

Cheung, V. K. M., Meyer, L., Friederici, A. D., & Koelsch, S. (2018). The right inferior gyrus processes nested non-local dependencies in music. *Scientific Reports, 8*.

Chiang, J. N., Rosenberg, M. H., Bufford, C. A., Stephens, D., Lysy, A., & Monti, M. M. (2018). The langage of music: Common neural codes for structured sequences in music and natural language. *Brain and Language, 185*, 30–37.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39*, Article e62.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181.

Clarke, E. F. (1987). Levels of structure in the organization of musical time. *Contemporary Music Review, 2*, 211–238.

Clay, E. R. (1882). *The alternative: A study in psychology* (2nd ed.). London: Macmillan.

Clewett, D., & Davachi, L. (2017). The ebb and flow of experience determines the temporal structure of memory. *Current Opinion in Behavioral Sciences, 17*, 186–193.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review, 121*, 33–65.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27*, 183–228.

Cona, G., & Semenza, C. (2017). Supplementary motor area as key structure for domain-general sequence processing: A unified account. *Neuroscience and Biobehavioral Reviews, 72*, 28–42.

Connolly, K., & Dalgleish, M. (1989). The emergence of a tool-using skill in infancy. *Developmental Psychology, 25*, 894–912.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception, 5*, 451–462.

Coull, J. T., Charras, P., Donadieu, M., Droit-Volet, S., & Vidal, F. (2015). SMA selectively codes the active accumulation of temporal, not spatial, magnitude. *Journal of Cognitive Neuroscience, 27*, 2281–2298.

Coupé, C., Oh, Y., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances, 5*, Article eaaw2594.

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford: Oxford University Press.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review, 24*, 1158–1170.

Craig, A. D. (2009). How do you feel - now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*, 59–70.

Craig, J. C., & Baihua, X. (1990). Temporal order and tactile patterns. *Perception and Psychophysics, 47*, 22–34.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 869–883.

Cunningham, D. W., Billock, V. A., & Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science, 12*, 532–535.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*, 255–267.

Daum, M. M., Prinz, W., & Aschersleben, G. (2009). Means-end behavior in young infants: The interplay of action perception and action production. *Infancy, 14*, 613–640.

Dawkins, R. (1976). Hierarchical organization: A candidate principle for ethology. In P. P. G. Bateson, & R. A. Hinde (Eds.), *Growing points in ethology*. Cambridge: Cambridge University Press.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970–974.

Deliege, I. (2001). Prototype effects in music listening: An empirical approach to the notion of imprint. *Music Perception, 18*, 371–407.

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

Di Lollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences, 16*, 317–321.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*, 481–507.

Dillon, B., Andrews, C., Rotello, C. M., & Wagers, M. (2019). A new argument for co-active parses during language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 1271–1286.

Durgin, F. H., & Sternberg, S. (2002). The time of consciousness and vice versa. *Consciousness and Cognition, 11*, 284–290.

Efron, R. (1963). The effect of handedness on the perception of simultaneity and temporal order. *Brain, 86*, 261–284.

Eimer, M., & Grubert, A. (2015). A dissociation between selective attention and conscious awareness in the representation of temporal order information. *Consciousness and Cognition, 35*, 274–281.

Ellis, R. J., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 264–280.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 229–240.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience, 9*, 878–879.

Fedorenko, E., McDermott, J., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the brain. *Journal of Neurophysiology, 108*, 3289–3300.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory and Cognition, 37*, 1–9.

Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences, 1369*, 132–153.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and WH-movement in german. *Brain and Language, 87*, 345–354.

Fink, M., Churan, J., & Wittmann, M. (2005). Assessment of auditory temporal-order thresholds - A comparison of different measurement procedures and the influences of age and gender. *Restorative Neurology and Neuroscience, 23*, 281–296.

Fink, M., Ulbrich, P., Churan, J., & Wittmann, M. (2006). Stimulus-dependent processing of temporal order. *Behavioral Processes, 71*, 344–352.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences, 1316*, 87–104.

Fiveash, A., McArthur, G., & Thompson, W. F. (2018). Syntactic and non-syntactic sources of interference by music on language processing. *Scientific Reports, 8*.

Fiveash, A., Thompson, W. F., Badcock, N. A., & McArthur, G. (2018). Syntactic processing in music and language: Effects of interrupting auditory streams with alternating timbres. *International Journal of Psychophysiology, 129*, 31–40.

Fougnie, D., Cormiea, S. M., & Alvarez, G. A. (2013). Object-based benefits without object-based representations. *Journal of Experimental Psychology: General, 142*, 621–626.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 35*, 1–36.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*, 244–249.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Friederici, A. D. (2019). Hierarchy processing in human neurobiology: How specific is it? *Philosophical Transactions of the Royal Society B, 375*.

Furstenau, N. (2014). Simulating bistable perception with interrupted ambiguous stimulus using self-oscillator dynamics with percept choice bifurcation. *Cognitive Processing, 15*, 467–490.

Genovesio, A., Tsujimoto, S., & Wise, S. P. (2009). Feature- and order-based timing representations in the frontal cortex. *Neuron, 63*, 254–266.

Gerstner, G. E., & Cianfarani, T. (1998). Temporal dynamics of human masticatory sequences. *Physiology and Behavior, 64*, 457–461.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Carnegie-Mellon University.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes, 8*, 147–161.

Gibson, E., & Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research, 29*, 231–240.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Goel, A., & Buonoamno, D. V. (2014). Timing as an intrinsic property of neural networks: Evidence from in vivo and in vitro experiments. *Philosophical Transactions of the Royal Society: B, 369*, 20120460.

Goodson, B. D., & Greenfield, P. M. (1975). The search for structural principles in children's manipulative play. *Child Development, 46*, 734–746.

Gottlieb, H., & Konecni, V. (1985). The effect of instrumentation, playing style, and structure in the Goldberg variations by Johann Sebastian Bach. *Music Perception, 3*, 87–102.

Goudar, V., & Buonomano, D. V. (2014). A model of order-selectivity based on dynamic changes in the balance of excitation and inhibition produced by short-term synaptic plasticity. *Journal of Neurophysiology, 113*, 509–523.

Granot, R. Y., & Jacoby, N. (2011). Musically puzzling I: Sensitivity to overall structure in the sonata form? *Musicae Scientiae, 15*, 365–386.

Greenfield, P. M. (1976). The grammar of action in cognitive development. In D. O. Walter, L. Rogers, & J. M. Finzi-Fried (Eds.), *Human brain function*. Los Angeles, CA: Brain Information Service, Brain Research Institute.

Greenfield, P. M., Nelson, K., & Saltzman, E. (1972). The development of rule-bound strategies for manipulating seriated cups: A parallel between action and grammar. *Cognitive Psychology, 3*, 291–310.

Guidali, G., Pisoni, A., Bolognini, N., & Papagno, C. (2019). Keeping order in the brain: The supramarginal gyrus and serial order in short-term memory. *Cortex, 119*, 89–99.

Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition, 2*, 45–78.

Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another's action. *Current Biology, 14*, 493–498.

Hansen, E. A., & Ohnstad, A. E. (2008). Evidence for a freely chosen pedalling rate during submaximal cycling to be a robust innate voluntary motor rhythm. *Experimental Brain Research, 186*, 365–373.

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*, 586–604.

Hardy, N. F., & Buonomano, D. (2016). Neurocomputational models of interval and pattern timing. *Current Opinion in Behavioral Sciences, 8*, 250–257.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19*, 304–313.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology, 6*(1962).

Herzog, M. H., Kammer, T., & Scharnowski, F. (2016). Time slices: What is the duration of a percept? *PLoS Biology, 14*, Article e1002433.

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences, 1*, 86–93.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience, 8*, 393–402.

Hilton, M., Räling, R., Wartenburger, I., & Elsner, B. (2019). Parallels in processing boundary cues in speech and action. *Frontiers in Psychology, 10*(1566).

Hirsh, I. J., & Sherrick, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology, 62*, 423–432.

Hoch, L., Poulin-Charronat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology, 2*(112).

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791–804.

Hogendoorn, H., Verstraten, F. A. J., & Johnston, A. (2010). Spatially localized time shifts of the perceptual stream. *Frontiers in Psychology, 1*.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences, 13*, 216–221.

Hommel, B. (2015). The theory of event coding (TEC) as embodied-cognition framework. *Frontiers in Psychology, 6*(1318).

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research, 1146*, 2–22.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception, 26*, 195–204.

Jacob, J., Breitmeyer, B. G., & Treviño, M. (2013). Tracking the first two seconds: Three stages of visual information processing? *Psychon. Bull. Rev., 20*, 1114–1119.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception and Psychophysics, 41*, 621–634.

Joseph, S., Kumar, S., Husain, M., & Griffiths, T. D. (2015). Auditory working memory for objects vs. features. *Frontiers in Neuroscience, 9*.

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 123–154.

Kahan, T. A., & Enns, J. T. (2014). Long-term memory representations influence perception before edges are assigned to objects. *Journal of Experimental Psychology. General, 143*, 566–574.

Kalm, K., & Norris, D. (2017). A shared representation of order between encoding and recognition in visual short-term memory. *NeuroImage, 155*, 138–146.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron, 53*, 427–438.

Karno, M., & Konecni, V. J. (1992). The effects of structural interventions in the first movement of Mozart's symphony in G-minor, K. 550, on aesthetic preference. *Music Perception, 10*, 63–72.

Khoei, M. A., Masson, G. S., & Perrinet, L. U. (2017). The flash-lag effect as a motion-based predictive shift. *PLoS Computational Biology, 13*, Article e1005068.

Kljajevic, V. (2010). Is syntactic working memory language specific? *Psihologija, 43*, 85–101.

Knoblich, G., Seigerschmidt, E., Flach, R., & Prinz, W. (2002). Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *Quarterly Journal of Experimental Psychology, 55A*, 1027–1046.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron, 50*, 963–974.

Koelsch, S. (2012). *Brain and music*. Hoboken, NJ: Wiley-Blackwell.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences, 110*, 15443–15448.

Konecni, V. (1984). Elusive effects of artists' messages. In W. R. Crozier, & A. J. Chapman (Eds.), *Cognitive processes in the perception of art*. New York: Elsevier.

Kopatich, R. D., Feller, D. P., Kurby, C. A., & Magliano, J. P. (2019). The role of character goals and changes in body position in the processing of events in visual narratives. *Cognitive Research: Principles and Implications, 4*.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C., & Kessler, E. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334–368.

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One, 10*, Article e0141069.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*, 72–79.

Lalitte, P., & Bigand, E. (2006). Music in the moment? Revisiting the effect of large scale structures. *Perceptual and Motor Skills, 103*, 811–828.

Lalitte, P., Bigand, E., Kantor-Martynuska, J., & Delbe, C. (2009). On listening to atonal variations of two piano sonatas by beethoven. *Music Perception, 26*, 223–234.

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Science, 10*, 494–501.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The hixon symposium* (pp. 112–146). New York: Wiley.

Lee, K. (2014). Perceptual reversal guided by integration between bottom-up input and top-down feedback over time course. *Psychologia, 57*, 12–30.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tine ratings. *Music Perception, 17*, 435–463.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford: Oxford University Press.

Lerdahl, F. (2015). Concepts and representations of musical hierarchies. *Music Perception, 33*, 83–95.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception, 24*, 329–336.

Levine, D., Hirsh-Pasek, K., Pace, A., & Golinkoff, R. M. (2017). A goal bias in action: the boundaries adults perceive in events align with sites of actor intent. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 916–927.

Levine, D., Buchsbaum, D., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Finding events in a continuous world: A developmental account. *Developmental Psychology, 61*, 376–389.

Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: additional evidence that auditory memory is absolute. *Attention, Perception and Psychophysics, 58*, 927–935.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.

Lewis, R. L. (2000). Faksifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research, 29*, 241–248.

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences, 1424*, 33–44.

Li, D., Cowan, N., & Saults, J. S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, and Psychophysics, 75*, 145–160.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249–336.

Liesefeld, H. R., & Müller, H. J. (2019). Current directions in visual working memory research: An introduction and emerging insights. *British Journal of Psychology, 110*, 193–206.

Lloyd, D. (2012). Neural correlates of temporality: Default mode variability and temporal awareness. *Consciousness and Cognition, 21*, 695–703.

Loaiza, V. M., & Camos, V. (2018). The role of semantic representations in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 863–881.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language, 31*, 713–733.

Loucks, J., Mutschler, C., & Meltzoff, A. N. (2017). Children's representation and imitation of events: How goal organization influences 3-year-old children's memory for action sequences. *Cognitive Science, 41*, 1904–1933.

Luo, Y. (2010). Do 8-month-old infants consider situational constraints when interpreting others' gaze as goal-directed action? *Infancy, 15*, 392–419.

MacDougall, H. G., & Moore, S. T. (2005). Marching to the beat of the same drummer: The spontaneous tempo of human locomotion. *Journal of Applied Physiology, 99*, 1164–1173.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*, 540–545.

Maffongelli, L., Antognini, K., & Daum, M. M. (2018). Syntactical regularities of action sequences in the infant brain: When structure matters. *Developmental Science, 21*, Article e12682.

Maffongelli, L., Bartoli, E., Sammler, D., Kölsch, S., Campus, S., Olivier, E., Fadiga, L., & D'Ausilio, A. (2015). Distinct brain signatures of content ans structure violation during action observation. *Neuropsychologia, 75*, 30–39.

Magliano, J., Kopp, K., McNerney, M. W., Radvansky, G. A., & Zacks, J. M. (2012). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition, 19*, 264–282.

Magliano, J., & Zacks, J. M. (2011). The impact of continuity editing in narrative film of event segmentation. *Cognitive Science, 35*, 1489–1517.

Makuuchi, M., Bahlmann, J., & Friederici, A. D. (2012). An approach to separating the levels of hierarchical structure building in language and mathematics. *Philosophical Transactions of the Royal Society B, 367*, 2033–2045.

Mallett, R., & Lewis-Peacock, J. A. (2018). Behavioral decoding of working memory items inside and outside the focus of attention. *Annals of the New York Academy of Sciences, 1424*, 256–267.

Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. *Brain and Language, 32*, 137–158.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multi-components view. *Neuropsychology, 8*, 506–523.

Martins, M. D., Gingras, B., Puig-Waldmueller, E., & Fitch, W. T. (2017). Cognitive representation of "musical fractals": Processing hierarchy and recursion in the auditory domain. *Cognition, 161*, 31–45.

Martins, M. D. J. D., Muršic, Z., Oh, J., & Fitch, W. T. (2015). Representing visual recursion does not require verbal or motor resources. *Cognitive Psychology, 77*, 20–41.

Martins, M. J. D., Bianco, R., Sammler, D., & Villringer, A. (2019). Recursion in action: An fMRI study on the generation of new hierarchical levels in motor sequences. *Human Brain Mapping, 40*, 2623–2638.

Martins, M. J. D., Fischmeister, F. P. S., Gingras, B., Bianco, R., Puig-Waldmueller, E., Villringer, A., Fitch, W. T., & Beisteiner, R. (2020). Recursive music elucidates neural mechanisms supporting the generation and detection of melodic hierarchies. *Brain Structure and Function*. https://doi.org/10.1007/s00429-020-02105-7

Mathias, S. R., & von Kriegstein, K. (2014). Percepts, not acoustic properties, are the units of auditory short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 445–450.

Matthews, N., & Welch, L. (2015). Left visual field attentional advantage in judging simultaneity and temporal order. *Journal of Vision, 15*(2). No. 7.

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience, 27*, 307–340.

McCabe, C. B., & Sanders, R. H. (2012). Kinematic differences between front crawl sprint and distance swimmers at a distance pace. *Journal of Sports Sciences, 30*, 601–608.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science, 366*, Article eaax0868.

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annual Review of Neuroscience, 36*, 313–336.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. Murray Hill, NJ: Prentice-Hall.

Michon, J. (1978). The making of the present: A tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 89–111). Hillsdale, NJ: Erlbaum.

Montemayor, C., & Wittmann, M. (2014). The varieties of presence: Hierarchical levels of temporal integration. *Timing and Time Perception, 2*, 325–338.

Nakai, T., & Okanoya, K. (2018). Neural evidence of cross-domain structural interaction between language and arithmetic. *Scientific Reports, 8*.

Nakai, T., & Sakai, K. L. (2014). Neural mechanisms underlying the computation of hierarchical tree structures in mathematics. *Ploe One, 9*, Article e111439.

Narmour, E. (1983). Some major theoretical problems concerning the concept of hierarchy in the analysis of tonal music. *Music Perception, 1*, 129–199.

Nee, D. E., & Jonides, J. (2013). Trisecting representational states in short-term memory. *Frontiers in Human Neuroscience, 7*.

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395*, 894–896.

Newtson, D., Engqvist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35*, 847–862.

Nicholls, M. E. R. (1994). Hemispheric asymmetries for temporal resolution: A signal detection analysis of threshold and bias. *Quarterly Journal of Experimental Psychology, 47A*, 291–310.

Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology, 12*, 359–368.

Nishikawa, N., Shimo, Y., Wada, M., Hattori, N., & Kitazawa, S. (2015). Effects of aging and idiopathic Parkinson's disease on tactile temporal order judgment. *PLoS One, 10* (3), Article e0118331.

Nishiyama, R. (2018). Separability of active semantic and phonological maintenance in verbal working memory. *PLoS One, 13*, Article e0193808.

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin, 143*, 992–1009.

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 948–957.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 411–421.

Oberauer, K. (2018). Removal of irrelevant information from working memory: Sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences, 1424*, 239–255.

Ögmen, 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*, Article e83671.

Ögmen, H., & Herzog, M. H. (2016). A new conceptualization of human visual sensory-memory. *Frontiers in Psychology, 7*.

Opitz, B., & Friederici, A. D. (2007). Neural basis of processing sequential and hierarchical syntactic structures. *Human Brain Mapping, 28*, 585–592.

Orbach, J., Ehrlich, D., & Heath, H. A. (1963). Reversibility of the necker cube: I. An examination of the concept of "satiation of orientation". *Perceptual and Motor Skills, 17*, 439–458.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences, 108*, 2522–2527.

Palmer, C., & Hutchins, S. (2006). What is musical prosody? *Psychology of Learning and Motivation, 46*, 245–278.

Palmer, C., Jungers, M., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45*, 526–545.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11*, 409–464.

Patel, A. D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience, 6*, 674–681.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.

Paton, J. J., & Buonomano, D. V. (2018). The neural basis of timing: Distributed mechanisms for diverse functions. *Neuron, 98*, 687–705.

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory and Cognition, 20*, 353–364.

Pelayo, P., Sidney, M., Kherif, T., Chollet, D., & Tourny, C. (1996). Stroking characteristics in freestyle swimming and relationships with anthropometric characteristics. *Journal of Applied Biomechanics, 12*, 197–206.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language, 87*, 539–558.

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B, 370*.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science, 306*, 499–503.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*, 229–259.

Pickering, M., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*, 1002–1044.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.

Po, J. M. C., Kieser, J. A., Gallo, L. M., Tésenyi, A. J., Herbison, P., & Farella, M. (2011). Time-frequency analysis of chewing activity in the natural environment. *Journal of Dental Research, 90*, 1206–1210.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*, 245–255.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience, 21*, 322–334.

Pöppel, E. (1978). Time perception. In R. Held, H. W. Leibowitx, & H. Teuber (Eds.), *Perception: 8. Handbook of sensory physiology* (pp. 713–729). Berlin: Springer-Verlag.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences, 1*, 56–61.

Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society B, 364*, 1887–1896.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*, 265–282.

Pozniak, C., Hemforth, B., & Scheepers, C. (2018). Cross-domain priming from mathematics to relative-clause attachment: A visual-world study in French. *Frontiers in Psychology, 9*(2056).

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*, 129–154.

Pritchett, B. L., Hoeflin, C., Koldeweyn, K., Dechter, E., & Fedorenko, E. (2018). High-level language processing regions are not engaged in action observation or imitations. *Journal of Neurophysiology, 120*, 2555–2570.

Prosser, S. (1995). Aspects of short-term auditory memory as revealed by a recognition task on multi-tone sequences. *Scandinavian Audiology, 24*, 247–253.

Protopapa, F., Hayashi, M., Kulashekhar, S., van der Zwaag, W., Battistela, G., Murray, M. M., Kanai, R., & Bueti, D. (2019). Chronotopic maps in human supplementary motor area. *PLoS Biology, 17*, Article e3000026.

Raburn, C. E., Merritt, K. J., & Dean, J. C. (2011). Preferred movement patterns during a simple bouncing task. *Journal of Experimental Biology, 214*, 3768–3774.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences, 17*, 133–140.

Rajendran, V. G., Teki, S., & Schnupp, J. W. H. (2018). Temporal processing in audition: Insights from music. *Neuroscience, 389*, 4–18.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review, 12*, 969–992.

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: Event models from perception to action. *Trends in Cognitive Sciences, 21*, 962–980.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music, 5*, 35–53.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosen, S. (1992). Temporal information is speech: Acoustic, auditory, and linguistic aspects. *Philosophical Transactions of the Royal Society: B, 336*, 367–373.

Rovelli, C. (2018). *The order of time*. London: Allen Lane.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437–442.

Sardroodian, M., Madeleine, P., Mora-Jensen, M. H., & Hanse, E. A. (2016). Characteristics of finger-tapping are not affected by heavy strength training. *Journal of Motor Behavior, 48*, 256–263.

Sardroodian, M., Madeleine, P., Voigt, M., & Hansen, E. A. (2014). Frequency and pattern of voluntary pedalling is influenced after one week of heavy strength training. *Human Movement Science, 36*, 58–69.

Sardroodian, M., Madeleine, P., Voigt, M., & Hansen, E. A. (2015). Freely chosen stride frequencies during walking and running are not correlated with freely chosen pedalling frequency and are insensitive to strength training. *Gait and Posture, 42*, 60–64.

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development, 8*, 113–128.

Scheepers, C., Galkina, A., Shtyrov, Y., & Myachykov, A. (2019). Hierarchical structure priming from mathematics to two- and three-site relative clause attachment. *Cognition, 189*, 155–166.

Scheepers, C., & Sturt, P. (2014). Bidirectional syntactic priming across cognitive domains: From arithmetic to language and back. *Quarterly Journal of Experimental Psychology, 67*, 1643–1654.

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science, 22*, 1319–1326.

Schell, M., Zaccarella, E., & Friederici, A. D. (2017). Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. *Cortex, 96*, 105–120.

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory and Cognition, 43*, 1021–1031.

Schellenberg, E. G., & Trehub, S. E. (2003). Accurate pitch memory is widespread. *Psychological Science, 14*, 262–266.

Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory and Cognition, 35*, 640–650.

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences, 1252*, 229–236.

Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, and Psychophysics, 80*, 1035–1056.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193–247.

Shivde, G., & Anderson, M. C. (2011). On the existence of semantic working memory: Evidence for direct semantic maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1342–1370.

Slevc, L. R., Rosenberg, J. A., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin and Review, 16*, 374–381.

Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology, 1*(175).

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1237–1256.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53*, 125–140.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(498).

Starke, S. D., & Baber, C. (2017a). Spontaneous bimanual independence during parallel tapping and sawing. *PLoS One, 12*, Article e0178188.

Starke, S. D., & Baber, C. (2017b). Movement consistency during repetitive tool use action. *PLoS One, 12*, Article e0173281.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*, 1169–1178.

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, 651–659.

Stevens, A. A., & Weaver, K. (2005). Auditory perceptual consolidation in early-onset blindness. *Neuropsychologia, 43*, 1901–1910.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)Consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645–659.

Szelag, E., von Steinbüchel, N., Reiser, M., de Langen, E. G., & Pöppel, E. (1996). Temporal constraints in processing of nonverbal rhythmic patterns. *Acta Neurobiologiae Experimentalis, 56*, 215–225.

Tadin, D., Lappin, J. S., Blake, R., & Glasser, D. M. (2010). High temporal precision for perceiving event offsets. *Vision Research, 50*, 1966–1971.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*, 427–440.

Tapia, E., & Beck, D. M. (2014). Probing feedforward and feedback contributions to awareness with visual masking and transcranial magnetic stimulation. *Frontiers in Psychology, 5*(1173).

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 252–260.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., & Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *The Journal of Neuroscience, 29*, 14726–14733.

Tillmann, B. (2012). Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science, 4*, 568–584.

Tillmann, B., & Bigand, E. (1996). Does formal musical structure affect perception of musical expressiveness? *Psychology of Music, 24*, 3–17.

Tillmann, B., & Bigand, E. (2001). Global context effect in normal and scrambled musical sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1185–1196.

Tillmann, B., & Bigand, E. (2004). The relative importance of local and global structures in music perception. *Journal of Aesthetics and Art Criticism, 62*, 211–222.

Tillmann, B., Bigand, E., & Madurell, F. (1998). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychological Research, 61*, 157–174.

Tipper, S. P. (2010). From observation to action simulation: The role of attention, eye-gaze, emotion, and body state. *Quarterly Journal of Experimental Psychology, 63*, 2081–2105.

Tsao, A., Sugar, J., Wang, C., Knierim, J. J., Moser, M.-B., & Moser, E. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature, 561*, 57–75.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830–846.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica, 116*, 185–203.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 683–692.

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology, 1*, 121–141.

Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition, 146*, 172–184.

Van de Cavey, J., Severens, E., & Hartsuiker, R. J. (2017). Shared structuring resources across domains: Double task effects from linguistic processing on the structural integration of pitch sequences. *Quarterly Journal of Experimental Psychology, 70*, 1633–1645.

van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research, 28*, 43–66.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*, 508–607.

VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences, 20*, 723–735.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences, 7*, 207–213.

VanRullen, R., Zoefel, B., & Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. *Philosophical Transactions of the Royal Society: B, Biological Sciences, 369*, 1–15.

Varley, R. A., Klessinger, N. J. C., Romanowski, C. A. J., & Siegal, M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences, 102*, 3519–3524.

Vollrath, M., Kazenwadel, J., & Krüger, H.-P. (1992). A universal constant in temporal segmentation of human speech. *Naturwissenschaften, 79*, 479–480.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1436–1451.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, and Psychophysics, 72*, 871–884.

Wernery, J., Atmanspacher, H., Kornmeier, J., Candia, V., Folkers, G., & Wittmann, M. (2015). Temporal processing in bistable perception of the necker cube. *Perception, 44*, 157–168.

White, P. A. (2017). The three-second "subjective present": A critical review and a new proposal. *Psychological Bulletin, 143*, 735–756.

White, P. A. (2018). Is conscious perception a series of discrete temporal frames? *Consciousness and Cognition, 60*, 98–126.

Williams, M., Hong, S. W., Kang, M.-S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bulletin and Review, 20*, 348–355.

Wingfield, A., & Butterworth, B. (1984). Running memory for sentences and parts of sentences: Syntactic parsin as a control function in working memory. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 351–364). Hillsdale, NJ: Erlbaum.

Winkler, M., Mueller, J. L., Friederici, A. D., & Männel, C. (2018). Infant cognition includes the potentially human-unique ability to encode embedding. *Science Advances, 4*, Article eaar8334.

Wittmann, M. (1999). Time perception and temporal processing levels of the brain. *Chronobiology International, 16*, 17–32.

Wittmann, M. (2009). The inner experience of time. *Philosophical Transactions of the Royal Society: B, Biological Sciences, 364*, 1955–1967.

Wittmann, M. (2011). Moments in time. *Frontiers in Integrative Neuroscience, 5*.

Wittmann, M. (2013). The inner sense of time: How the brain creates a representation of duration. *Nature Reviews Neuroscience, 14*, 217–223.

Wittmann, M., & Pöppel, E. (1999–2000). Temporal mechanisms of the brain as fundamentals of communication - with special reference to music perception and performance. *Musicae Scientiae, 3*(SI), 13–28.

Wittmann, M., von Steinbüchel, N., & Szelag, E. (2001). Hemispheric specialisation for self-paced motor sequences. *Cognitive Brain Research, 10*, 341–344.

Wittmann, M., Simmons, A. N., Flagan, T., Lane, S. D., Wackermann, J., & Paulus, M. P. (2011). Neural substrates of time perception and impulsivity. *Brain Research, 1406*, 43–58.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*, 1880–1882.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*, 979–1008.

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience, 23*, 4057–4066.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*, 273–293.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3–21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*, 29–58.

Zeng, T., Mao, W., & Liu, R. (2018). Structural priming from arithmetic to language in Chinese: Evidence from adults and children. *Quarterly Journal of Experimental Psychology, 71*, 1552–1560.

Zhang, J., Jiang, C., Zhou, L., & Yang, Y. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia, 91*, 490–498.