The Impure Phenomenology of Episodic Memory
Ali Boyle, University of Cambridge

Note: This is the submitted version of a paper to be published in Mind & Language. Citations should be to the final, published version.

Abstract: Episodic memory has a distinctive phenomenology: it involves ‘mentally reliving’ or ‘re-experiencing’ a past event. It’s been suggested that if episodic memory is characterized in terms of this phenomenology, then it will be ‘impossible to test’ for episodic memory in animals – because this is to characterize episodic memory in terms of its ‘purely phenomenological features’, which cannot be detected in animal behaviour. I argue that this is a mistake. The phenomenological features of episodic memory are impure phenomenological features – they have both subjective and objective aspects, and so can be detected in nonverbal behaviour. If animals’ memories exhibit these features, I argue, we should conclude that they have episodic memory. Insisting on a phenomenological characterization of episodic memory does nothing to damage the prospects for detecting it in nonhuman animals.

1. Introduction

All else being equal, scrub-jays prefer worms to nuts. This has been exploited in a series of experiments investigating their memory capacities. In one experiment, scrub-jays are given the opportunity to cache worms and nuts at different sites. Later, they have the opportunity to return to these caches, the food having been removed and the sand replaced to eliminate visual and olfactory cues. All else equal, they tend to return to the cache with the worms. One respect in which things are unequal is that worms decay at a faster rate, quickly becoming inedible. Scrub-jays are sensitive to this: if the worms have been cached long enough to decay, the scrub-jays will tend to return to the nuts instead. This is taken to show that scrub-jays remember not just what is buried where, but also when it was buried (Clayton & Dickinson, 1998).

Episodic memory is the form of declarative memory enabling one to recall events from one’s ‘personal past’ – events experienced first-hand. It’s typically contrasted with semantic memory – memory for facts. When Endel Tulving (1972) first described episodic memory, he characterized it as the ability to recall what happened where and when. On this way of thinking about episodic memory, the experiment just described provides evidence that scrub-jays have episodic memory – so it is not, as some have suggested (Suddendorf & Corballis, 2007), uniquely human.

1 [Acknowledgment removed.]
This ‘what-where-when’ view of episodic memory has been challenged, though. Its critics claim that it fails to distinguish episodic memory from semantic memory – since semantic memory can also record the what-where-when of a past event (S. B. Klein, 2013; Suddendorf & Busby, 2003). What really distinguishes episodic and semantic memory is not that episodic memory has this content, but that episodic memory essentially involves an experience of ‘mentally reliving’ the remembered event, which is absent in semantic memory. We may remember when and where we were born – but these memories are semantic, not episodic, since (mercifully) we cannot ‘mentally relive’ our births. Characterizing episodic memory in terms of ‘what-where-when’ content consequently neglects episodic memory’s most essential feature. So, the thought goes, we should instead characterize episodic memory in terms of its distinctive phenomenology: to episodically remember an event is to have an experience of ‘mentally reliving’ it (Tulving, 2005). If this is what episodic memory is, though, the experiment just described does not bear on whether nonhuman animals have episodic memory.

The problem is that this renders it obscure what sort of evidence could bear on that question. We know that humans have experiences of ‘mentally reliving’ because they tell us so – but animals are not in a position to report on the felt qualities of their experiences. For this reason, a phenomenological characterization of episodic memory is widely agreed to be useless, when it comes to investigating episodic memory in animals. As critics of phenomenological characterizations point out, determining whether animals have episodic memory requires characterizing it in terms of its ‘objectively defined features as opposed to purely phenomenological ones, such as the type of information encoded’ (Clayton, Russell, & Dickinson, 2009). Since a phenomenological characterization fails to do this, it is ‘straightforwardly inapplicable’ to animal behaviour (Michaelian & Sutton, 2017); it makes it ‘impossible to test [for episodic memory] in animals’ (Clayton & Dickinson, 2010; see also Allen & Fortin, 2013; Eacott & Easton, 2012). For this reason, many researchers proceed by endorsing a variant on the ‘what-where-when’ account but labelling the sort of memory it describes ‘episodic-like’ memory – where ‘episodic-like’ is intended to indicate agnosticism about whether the form of memory in question is ‘accompanied by conscious recollection’ (Clayton, Salwiczek, & Dickinson, 2007).²

² This variant adds that what-where-when information is integrated into a single representation which can be deployed flexibly (Clayton, Bussey, & Dickinson, 2003). The flexibility condition marks episodic memory as a type of declarative memory. The integration condition entails that
My goal in this paper is to show that this is a mistake: characterizing episodic memory in terms of its phenomenology does not make it impossible to test for in non-human animals. There is consequently no need for scientists to limit their ambition to the detection of ‘episodic-like’ memory. The mistake here is in thinking that a phenomenological characterization of episodic memory must characterize it in terms of ‘purely phenomenological features’, rather than in terms of ‘objective’ ones. I take the objective features of a cognitive state or progress to be ones which could show up in behaviour – ones which can be ‘behaviourally defined’ (Clayton & Dickinson, 2010). These include its representational features, and its functional role. I take ‘purely phenomenological features’, on the other hand, to denote phenomenological features which cannot be so defined, because they do not show up in behaviour – the non-representational, non-functional felt qualities of experience which philosophers call ‘qualia’. In short, the idea underlying this criticism of the phenomenological view is that the phenomenological features of episodic memory are representationally and functionally impotent.

I argue that this is not the case for any of the distinctive phenomenological features of episodic memory. I begin in §2 by characterizing the phenomenology of episodic memory in terms of five distinctive phenomenological features. In §3, I argue that each of these five features is an ‘impure phenomenological feature’ – a term I introduce to denote a phenomenological feature which makes a functional or representational difference. These impure phenomenological features, I argue in §4, can be detected in non-verbal behaviour. So, it is possible to both characterize episodic memory in terms of its distinctive phenomenology and detect it in nonhuman animals.

2. Five Phenomenological Features

A phenomenological characterization of episodic memory takes the phenomenology of episodic memory to be among its defining features – such that to ask whether an individual has episodic memory is to ask whether they have this phenomenology. There is no canonical account of what the relevant phenomenology is, however. Typically, it is glossed using phrases if the subject retrieves one element of the ‘what-where-when’ content, she retrieves them all. It is not clear that this helps in distinguishing episodic and semantic memory. First, it is unclear how integration should be measured (Zentall, 2005); second, examples of episodic memories in which one or more of the ‘what-where-when’ elements is missing are readily available (Suddendorf & Busby, 2003).
like ‘mentally reliving’, ‘re-experiencing’ or ‘replaying in the mind’s eye’, which are all somewhat metaphorical. The purpose of this section is to characterize this phenomenology more concretely.

One thought naturally suggested by these metaphors is that an experience of episodically remembering is in some way phenomenologically similar to one’s experience at the time of the remembered event. But this does not seem to exhaust the distinctive phenomenology of episodic recall. Tulving (2005) highlights two further aspects to this phenomenology: that it involves a kind of self-consciousness he calls ‘autonoesis’, and an awareness of time he calls ‘chronesthesia’. But whilst it is widely accepted that episodic memory involves time- and self-consciousness, precisely what this amounts to is unclear. In what follows, I explore the sense in which episodic memory involves phenomenological similarity, time-consciousness and self-consciousness, and extract five phenomenological features of episodic memory.

First, phenomenological similarity. Whilst there is something intuitively correct about the idea that having an episodic memory feels similar to one’s experience at the time of the remembered event, this claim needs some qualification if it is to be plausible. Importantly, it cannot be the claim that one’s experience in episodically remembering is phenomenologically identical to one’s original experience, for two reasons.

Firstly, the constructive processes underpinning episodic memory ensure that, frequently, what is remembered differs from what was experienced (Michaelian, 2015). For instance, memories are often experienced in ‘observer perspective’, rather than ‘field perspective’ – from the point of view of a third party looking on, rather than from one’s own perspective at the time of the event. So, when remembering an event in observer perspective, one might be able to ‘see’ oneself (Nigro & Neisser, 1983). There are good reasons for thinking that these observer perspective memories are genuine episodic memories (Debus, 2007; Michaelian, 2015). It follows from this that episodic memories can differ in content from one’s original experience – since one cannot possibly have seen oneself from a third-person perspective at the time of the event – and that there is a corresponding difference in phenomenology. So, despite its phenomenological similarity to the original experience, an episodic memory may differ representationally and phenomenologically from one’s experience of the remembered event.

Secondly, although there are phenomenological similarities between seeing an F and mentally reliving seeing an F, there are significant differences. It would be difficult to mistake
an experience of the first kind for one of the second, even if the ‘mental reliving’ had not been altered by constructive processes. A key difference is that perceptual experiences have, and relivings lack, ‘presence’ – the sense that the experience’s character is directly responsive to the character of its objects at the time of the experience (Crane & French, 2017). So, we should not say that episodic memory literally has perceptual phenomenology, although it involves experiences which ‘characteristically correspond to our use of the five senses’ (Martin, 2001; see also Debus, 2007). Instead, episodic memory involves ‘quasi-perceptual’ phenomenology – experiences which are similar to ones of seeing, hearing etc., but which differ in respect of presence. To put it another way, episodic memory involves what is often called ‘mental imagery’ – especially visual mental imagery (Hoerl, 2001). These points together suggest that episodic memories are phenomenologically similar to experiences of remembered events in that they involve mental imagery which represents those events. So, we can extract the following phenomenological feature of episodic memory:

(1) Mental Imagery: episodic memory involves mental imagery representing the remembered event.

Next, episodic memory is widely held to involve a sort of time-consciousness – such that to lack episodic memory is to be ‘mentally stuck in time’ (Roberts, 2002). As noted above, Tulving has coined the term ‘chronesthesia’ to refer to the sort of time-consciousness involved in episodic memory - but the meaning of this term is obscure. At one point, Tulving (2005) characterizes chronesthesia in terms of ‘attaching a temporal marker’ to a memory; at another, as involving an ‘ever-present awareness of one’s being existing in a subjective sea of time, always in transition from what is now becoming the past to what was once the future’. These ideas are non-equivalent – and as Michaelian (2015) notes, the second seems to describe the sort of temporal awareness present in ordinary unfolding experience, rather than anything distinctive of episodic memory. Moreover, Tulving frequently characterizes chronesthesia as an awareness of ‘subjective time’ – something ‘related to but not identical with physical time’ (2005). Subjective time is variously glossed as the time ‘in which we live’, the time through which we ‘mentally travel’, and ‘the thought-about time in which one’s personal experiences take place’ (Tulving, 2002). But none of these describe a subjective temporal dimension. Most obviously, we live in physical time. But the time in which we ‘mentally travel’ is also physical
time: in episodically remembering, we mentally revisit some moment in the (physical) past.³ Relatedly, whilst memories themselves unfold in the present, the ‘thought about’ time is the physical past. Consequently, the terms ‘chronesthesia’ and ‘subjective time’ seem to add unnecessary obscurity. More importantly, introducing a term to refer to ‘the’ form of time-consciousness involved in episodic memory masks the possibility that it involves time-consciousness in more than one way – which, I suggest now, is the case.

One difference between episodic and semantic memories is that the former have a felt duration and temporal structure: they seem to unfold over time. Suppose I episodically remember a time when I came home to a dark house, turned on the light, and was greeted by a surprise party. The memory is not presented all at once, but unfolds in a particular way: first, I ‘relive’ coming home, then hitting the lights, then seeing the guests. This marks a phenomenological contrast between episodic and semantic memory. Semantic recall occurs in time, like everything else. But in semantic recall, it may appear to the subject that the memory is given to her ‘all at once’, rather than seeming to unfold. This suggests a second phenomenological feature of episodic memory:

(2) Temporal Structure: Episodic memory has apparent temporal structure.

In addition to this, episodic memories seem to be accompanied by a sense of ‘pastness’. When we episodically remember, it seems as though the episode being relived ‘belongs’ to the past (Clayton et al., 2007; Tulving, 1984). This feature is distinct and dissociable from having an apparent temporal structure. Earworms, for instance – experiences of song fragments or musical phrases repeatedly ‘playing’ in the mind – appear to unfold, but do not seem to belong to the past. By contrast, the thought goes, when we episodically remember an event, it seems to us that what we are remembering is something which happened at some moment in the past. This gives us the third phenomenological feature of episodic memory:

³ This is not to say that we literally travel back in time when we remember. But there is no need to account for this by introducing non-physical time; we do not travel in time because we do not literally ‘travel’ at all. Nor should we be persuaded to introduce subjective time on the grounds that ‘past’ and ‘future’ are ‘necessarily defined with respect to a sentient observer’ (Nyberg, Kim, Habib, Levine, & Tulving, 2010). ‘Here’ and ‘there’ are also observer relative notions, but this does not motivate postulating ‘subjective space’ to account for our ability to remember being elsewhere.
(3) Sense of Pastness: episodic memory is associated with a 'sense of pastness'.

Finally, in addition to involving time-consciousness and mental imagery, episodic memory is widely agreed to involve self-consciousness. Once again, Tulving has coined a term to refer to the type of self-consciousness in question – ‘autonoesis’. But again, the term obscures that there are at least two ways in which episodic memory is a self-conscious affair. Typically, ‘autonoesis’ is introduced by appeal to the authority of William James – in particular to two claims he made about memory. But plausibly, each of these claims highlights a distinct variety of self-consciousness.

First, James (1890) wrote that memory ‘requires more than the mere dating of a fact in the past. It must be dated in my past’. A natural interpretation of this claim is that episodically remembered events are presented as ones belonging to my ‘personal past’ (Michaelian, 2015) – my personal timeline, comprising those events in which I was involved, or which happened to me. This suggests a fourth phenomenological feature of episodic memory:

(4) Self-Involvement: episodic memory presents events as self-involving.

James also wrote that memories are accompanied by a feeling of ‘warmth and intimacy’ (1890). To see what this amounts to, it is helpful to introduce a clinical case often construed as one in which this feeling of ‘warmth and intimacy’ went missing from the patient’s episodic memories. The patient, RB, suffered a head trauma following which he was able to remember particular incidents from his past, complete with ‘temporal, spatial and self-referential’ information, but ‘did not feel that the memories he experienced belonged to him’ (S. B. Klein & Nichols, 2012). In other words, he lacked a ‘sense of ownership’ over his memories. For instance, of one memory he said

‘I can picture the scene perfectly clearly […] I can ‘relive’ it in the sense of re-running the experience of being there. But it has the feeling of imagining [as if] re-running an experience that my parents described from their college days. […]

Intellectually, I suppose I never doubted that it was a part of my life.’

(S. B. Klein & Nichols, 2012)
RB was able to remember past events and his own involvement in them — so, plausibly, he was able to date the remembered events in his past. But, on the standard interpretation of the case, there was something missing from his memories — the ‘warmth and intimacy’, in virtue of which we feel that our memories ‘belong’ to us. This suggests a final phenomenological feature of episodic memory:

(5) Sense of Ownership: episodic memory is (in non-pathological cases) accompanied by a sense of ownership.

A phenomenological characterization of episodic memory claims that it essentially involves an experience of ‘mentally reliving’. ‘Mental reliving’, I have argued, comprises five distinctive phenomenological features: (1) Mental Imagery; (2) Temporal Structure; (3) Sense of Pastness; (4) Self-Involvement; (5) Sense of Ownership. One might object to such a characterization by denying that ‘mentally reliving’ actually involves all of (1)-(5), or by denying that (1)-(5) are really phenomenological features. These are legitimate challenges, but I will set them aside. I am aware of nobody denying that episodic memory has these features. My primary concern in what follows is to show that characterizing episodic memory in this way does not preclude detecting it in animals. I will assume that these features really are phenomenological ones — but if they are not, this is simply grist to my mill. My goal is to show that these are not purely phenomenological features; they cannot possibly be purely phenomenological if they are not phenomenological at all.

3. Impure Phenomenological Features

I will now argue that each of the phenomenological features (1)-(5) outlined in §2 is an ‘impure phenomenological feature’. An impure phenomenological feature is a phenomenological feature which ‘does something for us’. More precisely: some feature of a mental process is an impure phenomenological feature iff (a) it contributes to the felt quality of the process and (b) it contributes to the functional or representational features of the process.

As an example of an impure phenomenological feature, consider the felt unpleasantness of pain. This obviously contributes to pain’s overall phenomenology. It also appears to contribute to its functional role: it is because pain feels unpleasant that it motivates

4 [Acknowledgment removed for anonymous review.]
us to behave in certain ways. We consequently find it reasonable to take certain behaviours to be evidence of pain. For instance, motivational trade-off, in which an animal appears to trade-off its preference for avoiding a harmful stimulus against its other preferences, is taken to be a ‘credible indicator’ of pain (Birch, 2017) because the hypothesis that the animal is in pain provides the best explanation of its behaviour (Tye, 2017). The unpleasantness motivates the animal to avoid the harmful stimulus, at the cost of foregoing other goods. Conversely, individuals with pain asymbolia report feeling pain, but are not motivated to stop it. On one interpretation of the condition (Grahek, 2011), these individuals feel ‘pain without unpleasantness’ – they feel some aspects of pain, but not its unpleasantness. Because of this, they are not motivated to avoid or stop pains. Felt unpleasantness, then, does something for us: its presence or absence makes a difference to pain’s functional role.

This is not to say that this felt unpleasantness is not conceivably separable from pain’s functional features; it is. But I am not concerned here with metaphysical possibilities, but with how things actually are. As a matter of fact, perhaps of nomological necessity, the unpleasantness of pain seems to contribute both to the felt quality of pain and to its functional role. The felt quality explains why individuals in pain behave as they do; where that felt quality is missing, they behave differently. So, although we can conceptually isolate the felt quality of unpleasantness from the role it plays in our mental economy, unpleasantness is nevertheless an impure phenomenological feature of pain.

So too, the phenomenological features listed in §2 are impure phenomenological features of episodic memory. Each of them does something for us, contributing something to the representational and functional features of our episodic memories. To show this, I consider each of these features in turn.

First, mental imagery. That episodic memory involves mental imagery contributes something to its phenomenal character. But it is implausible on its face to take this to be a ‘purely’ phenomenological feature of episodic memory – since mental imagery has distinctive representational features, in addition to its distinctive phenomenology. What these representational features are is a matter of debate. But whatever they are, if episodic memory involves mental imagery, these are representational features of episodic memory too. So mental imagery is one of episodic memory’s impure phenomenological features.

---

5 Not the only one (see C. Klein, 2015).
Of course, this is unhelpful without knowing what the representational features of mental imagery are. But without entering into controversies about the nature of mental imagery, we can make a number of plausible claims about its representational features. I focus primarily on the case of visual mental imagery – though what I say should generalise. As a first pass, we can say that visual mental imagery represents the visible features of objects. So, when we have visual mental imagery of a chair, we represent its visible and spatial features – it has a certain shape, size, orientation – a left and a right – etc. If our imagery represents a number of objects, then we will represent the spatial relations between these objects.

Whether or not mental imagery represents using a genuinely image-like format, there is a similarity worth noting between mental imagery and pictorial representation. Pictures involve a kind of mandatory commitment: they must represent somewhat determinately the visible properties of the things they represent. As Sober (1976) writes, ‘there are concepts that cannot stand alone in the interpretation of pictures’. No picture, for instance, represents ‘that something is triangular full stop’ – it must represent some specific kind of triangle. Similarly, a picture cannot represent simply that two people are in a room – it must represent that they are sitting or standing, that they are in some part of the room and stand in some spatial relation.

Mental imagery plausibly also involves mandatory commitment. To have a mental image of a man in a room is to have an experience representing a man in a room with quasi-visual phenomenology – one which feels somewhat like seeing a man in a room. But there’s nothing it’s like to see a man in a room, but not to see him as having any particular size, shape, colour, location and so on. A visual experience must represent these visuo-spatial features somewhat determinately. A visual mental image must do the same, otherwise it is difficult to see how it might have anything like visual phenomenology. This is not to say that visuo-spatial features must be represented in a fully determinate way. Visual experiences and mental imagery admit of degrees in resolution, just as pictures do (Nanay, 2015; Sober, 1976). Nevertheless, it seems likely that there is a lower bound on the level of specificity and detail required in any experience with visual, or quasi-visual, phenomenology.

Given that in episodic memory, we recall events which might involve a number of individuals or objects interacting in the context of a shared spatial environment, episodic

---

6 That episodic memories involve mandatory commitment may go some way toward explaining boundary extension, wherein individuals report remembering more than they saw – for instance, remembering as whole objects they saw partly obscured (Intraub, Bender, & Mangels, 1992).
memories must be rich in visuo-spatial detail. To construct a quasi-visual experience of an event like giving a lecture requires representing somewhat determinately the visuo-spatial features of the room, the audience and so on. In short, the sorts of things episodic memories are memories of – the perceptible features of objects interacting in shared spatial environments – cannot ‘stand alone’ in mental imagery, but must be represented somewhat determinately. If this is right, then from the fact that episodic memory involves mental imagery, we can conclude that it represents in some detail the visuo-spatial features of the remembered events.

The idea that episodic memory stores information about the visuo-spatial features of events is hardly new. Tulving (1972), for instance, suggests that in episodic memory, events are represented ‘in terms of (a) its perceptible properties, and (b) its temporal-spatial relation to other experienced events’. And a more recent ‘minimalist’ account of episodic memory, developed by James Russell and colleagues (Clayton & Russell, 2009; Russell & Hanna, 2012), emphasises the spatial content of episodic memory – this time on Kantian grounds.7 What I am suggesting, though, is that episodic memory represents visuo-spatial features precisely because it involves mental imagery – because it is in the nature of mental imagery to have such content. So, mental imagery is an impure phenomenological feature of episodic memory: as well as contributing to its phenomenology, it makes a difference to its representational features.

Similarly, the apparent temporal structure of episodic memory contributes to its representational features: just as episodic memories represent the visuo-spatial features of events through mental imagery, they represent events’ temporal features through their apparent temporal structure. In fact, it is helpful to characterize the way in which memories appear to unfold in terms of mental imagery. Episodic memories seem to unfold in the sense that the mental imagery it involves is dynamic: it changes over time in a way that reflects how the perceptible features of the remembered event changed over time. In this sense, and whilst there are many disanalogies between episodic memories and videos, the unfolding of an episodic memory seems to be representationally on a par with the unfolding of a video. An episodic memory in which one first ‘relives’ X and then Y represents that X occurred before Y; if it unfolded in the reverse order, it would represent that Y happened before X.

---

7 Their thought is that since perceptual experience, according to Kant, necessarily has an egocentric spatial structure, and episodic memory involves ‘re-experience’, episodic memory must inherit the egocentric spatial structure of the original experience. This account interprets ‘re-experience’ in the implausibly literal sense I rejected above (Hoerl, 2014), and in insisting that episodic memory has egocentric content, excludes observer perspective memories from the category of episodic memory.
The apparent temporal structure of an episodic memory, then, should be counted among its impure phenomenological features—since it is by means of this apparent structure that episodic memories represent the temporal structure of events. That episodic memory involves representing temporal order is emphasised in Russell and colleagues’ minimalist account of episodic memory, and has motivated studies of episodic memory in both rodents (Allen, Morris, Mattfeld, Stark, & Fortin, 2014) and children (Burns, Russell, & Russell, 2015). But precisely the reason for thinking that episodic memory represents temporal order is that this is what the apparent temporal structure transparently represents. So, apparent temporal structure is an impure phenomenological feature of episodic memory.

Third, the ‘sense of pastness’. This is often explicitly characterized in representational terms—as a matter of attaching a ‘temporal marker’ to an event (Tulving, 2005), or representing when it happened (Michaelian, 2015). But it is doubtful that episodic memory constitutively involves representing when an event happened: it is not uncommon to recall a past event without representing when it occurred (Russell & Hanna, 2012; Suddendorf & Busby, 2003). Nevertheless, unlike imagined events, remembered events do not seem to be free of the temporal order—they still seem to ‘belong’ to the past, even when the subject does not know where in the past they belong. So, the feeling of pastness is not simply a matter of representing when something happened.

Hoerl (2008) suggests that the sense of pastness consists in having a kind of ‘tensed thought’. Episodic memories do not represent the present state of the world; they represent past states of the world as such. To represent a past event as past is to represent it as standing in a certain causal relationship to the present. Hoerl (2014) characterizes this in terms of appreciating that the event has been ‘superseded by other objects or situation’—and so cannot be revisited or repeated (see also Russell & Hanna, 2012). This does seem to be part of what is involved in representing something as past. Besides this, another important causal connection between past and present events is this: that an event occurred in the past has a certain causal and practical relevance for what happens now and in the future. That a particular event occurred may raise the probability of another event occurring, and it may make it rational to do certain things now or in the future. Representing a past event as such involves appreciating this— that despite having been superseded by the present, the event is causally and practically relevant to present and future events.

So understood, the sense of pastness is an impure phenomenological feature. For episodic memory to have the ‘sense of pastness’ is not only for it to have a felt quality, but for
it to represent past events as such. As a result of constitutively involving this sort of representation, the sense of pastness contributes to the functional role of episodic memory – because it is only in virtue of representing past events as past that episodic memories can rationally play the role that they do in practical reasoning. Suppose that I episodically remember my friend Henry heading into Sally’s party, and I have just learned that a co-worker contracted chicken pox at that same party. My memory, together with what I have just learned and my desire not to get chicken pox, ought to motivate me to avoid Henry until I find out whether he is infected. My memory plays this role in my practical reasoning only because it represents the event – Henry heading into the party – as a past event. If it did not represent this event as one which had actually occurred, but as belonging to some imagined or future time, it could not rationally motivate me to avoid Henry. So, the sense of pastness, is doubly impure: it contributes to both the representational and functional features of episodic memory.

The fourth feature, self-involvement, is a relatively clear-cut case. For an episodic memory to present an event as one in which I was involved, or for which I was at least present, seems straightforwardly to imply that it must have some self-referential content. It is hard to see how a memory could present an event as self-involving without having such content – without representing something about one’s own involvement in or presence at an event. Consequently, self-involvement is not a purely phenomenological feature of episodic memory.8

Finally, the sense of ownership is often glossed in metarepresentational terms. Tulving (2005) characterizes it in terms of the subject’s representing ‘that one had a particular experience […] in a particular place at a particular time’; Michaelian (2015) suggests that it involves ‘representing an experience of that [remembered] event that belongs to me’. This metarepresentational construal gels, in some cases, with RB’s introspective reports. After his accident, he said that his memories felt like events that had been ‘described by someone else’ – but once he recovered, they seemed like things he ‘had done and experienced’ (S. B. Klein & Nichols, 2012).

But a simple metarepresentational construal does not really capture what was distinctive about RB’s memories. RB’s episodic memories may well have represented that he had certain experiences – just as my memory of a friend’s wedding may represent that she witnessed certain things or was happy. What distinguished RB’s memories from typical

---

8 One might wonder whether this is even a phenomenological feature – but as noted, this is unimportant for my purposes.
episodic memories is that they did not seem, from his perspective, to have their *source* in experience. RB’s memory may have represented that he witnessed an event, but it did not seem to him that he had this memory for that reason: it was as though he had this memory because the event was described to him. This suggests that what was missing from RB’s memories was not (simply) metarepresentational content, but *source information*.

In this vein, Jerome Dokic (2014) suggests that episodic memory involves a feeling that the memory is ‘first-hand’, i.e. that it ‘originates directly in our own experience’. Similarly, Michaelian (2015) proposes that the sense of ownership might be thought of as the result of ‘process monitoring’ mechanisms – signalling that the memory is the result of a memory process, rather than an imaginative one. And Johannes Mahr and Gergely Csibra (2018) suggest that ‘autonoesis marks those events of which one had first-hand experience as opposed to some other source’. There are differences between these views, but the broad suggestion is that the sense of ownership is the product of source monitoring, and enables a subject to discriminate between first- and second-hand information – in order to ‘reduce uncertainty about whether the remembered events actually occurred’ (Michaelian, 2015). If this is right then the sense of ownership, like the sense of pastness, modulates the role episodic memories play in practical reasoning. The absence of this feeling may decrease the likelihood of the subject endorsing and acting upon the memory. Consequently, the sense of ownership is an impure phenomenological feature of episodic memory: its presence or absence has significant consequences for the role that a memory plays in practical reasoning and the production of behaviour.

This demonstrates that the sense of ownership is an impure phenomenological feature. It also calls into question whether we should expect episodic memory to be accompanied by a sense of ownership in all species. If the sense of ownership plays a ‘source monitoring’ role, it seems unlikely that animals would have any use for it. The utility of source monitoring in the case of human memory is clear. Humans receive information through both testimony and experience, and construct subjectively similar memory representations on the basis of both sorts of information (Loftus, 2005). This means it is not uncommon for us to have states which are subjectively like episodic memories, but whose content derives from a source other than our experience of an event. But animals, presumably, do not receive information through testimony, or any source other than experience, so a source monitoring mechanism would be
of little use to them. This, together with the suggestion from RB’s case that episodic memory and the sense of ownership are dissociable, makes it reasonable to think that if there is nonhuman episodic memory, it may not have an associated sense of ownership – so, this should not be treated as an essential feature of episodic memory, if our goal is to detect episodic memory in animals.

I have argued that each of the phenomenological features of episodic memory highlighted in §2 is an impure phenomenological feature. That is to say, each of these features contributes to the representational and functional features of episodic memory, as well as to its phenomenology. So, each of these features has both an objective and a subjective aspect. Mental imagery and apparent temporal structure are the means by which episodic memories represent the visuo-spatial and temporal features of events. The sense of pastness constitutively involves representing a past event as such, grounding episodic memory’s role in practical reasoning. For a memory to present events as self-involving is for it to have some self-referential content. And the sense of ownership, where it exists, functions to provide the subject with source information.

As in the case of the unpleasantness of pain, the subjective and objective aspects of (1)-(5) conceivably come apart. We can imagine a kind of ‘zombie memory’ having all the objective features I have described but lacking any of episodic memory’s phenomenological features. So, one might be tempted here to accuse me of sleight of hand – to say that there are purely phenomenological features in the vicinity, namely the felt qualities missing in this thought experiment, which I have simply run together with some contingently associated objective features. If this is right, we cannot (as I will propose in §4) investigate whether animals have episodic memory by investigating whether their memories exhibit the objective aspects of (1)-(5). Even if they do, the objection might go, it will be a further question whether the relevant phenomenology is present.

This objection misses the mark, since the question at hand – whether animals have episodic memory – concerns only what is actually the case. Answering this question requires only addressing empirical possibilities, not merely metaphysical ones. But there is no reason for supposing that the phenomenological aspects of (1)-(5) come apart from their objective

---

9 Note that when Jonathon Crystal and colleagues (2014; 2013) claim that rats have source memory, they are not referring to source monitoring as I use the term here. Their studies suggest that rats recall features of an event besides what was where and can discriminate events on that basis – but not that they take their memories to originate in experience as opposed to another source.
aspects, or that zombie memory is an empirical possibility. Quite the opposite: I have been arguing that the objective features of episodic memory go hand in hand with its phenomenology. So, whilst we can in thought dissociate the felt qualities of (1)-(5) from their contribution to the functional and representational features of episodic memory, we should doubt that they are in fact so dissociable. Consequently, it is not sleight of hand to insist that these are impure phenomenological features. Here, as in other areas where objective features and phenomenology march in step, functional equivalence should be regarded as evidence of phenomenology.

4. Behavioural Detection

‘Mental reliving’ comprises five phenomenological features, which I have argued are impure phenomenological features. So, in characterizing episodic memory in terms of the phenomenology of mental reliving, we indirectly highlight a number of objective features of episodic memory. In particular (and disregarding the sense of ownership for the reasons outlined in §3), episodic memory:

(A) stores detailed information about visuo-spatial features of an event
(B) represents the event’s temporal structure
(C) represents a past event as past
(D) stores some self-specifying information about the subject at the time of the event

My proposal, then, is that we investigate whether animals have episodic memory by investigating whether they have memories exhibiting these features. This proposal can be implemented only if these features can be detected in nonverbal behaviour. In this section, therefore, I make some suggestions about how this might be done.

First, though, note that in making this proposal, I do not mean to imply that (A)-(D) exhaust the nature of episodic memory. (A)-(D) are simply the objective features which fall out when we characterize episodic memory in terms of its phenomenology. There is consequently an evidential connection between having memories with these features and having episodic memories. Most obviously, if an individual does not have a memory of an event with features (A)-(D), this indicates that she does not have episodic memory. Conversely, I suggest, if a person does have a memory of an event with features (A)-(D), this should be counted as evidence that she episodically remembers it.
One might take issue with my claim that this evidential connection holds in the second direction, as well as the first. To see why, recall that the challenge to the ‘what-where-when’ view of episodic memory was that semantic memory can encode what-where-when content, just as episodic memory can. So, evidence that animals recall what-where-when content does not support the episodic memory hypothesis. In this respect, one might object that the what-where-when view and the phenomenological view, as I have presented it, are companions in guilt. I have suggested that episodic memories have features (A)-(D). But, the objection goes, a semantic memory could have these features too: a semantic memory might store detailed information about a past event’s visuo-spatial and temporal features, and self-specifying information about one’s own involvement in the event. So, if animals could retrieve such information, this would not indicate that they had episodic memory. If their memories lack these features, that may well count against their being episodic – but their exhibiting these features provides no reason to favour the hypothesis that they have episodic memory.

A semantic memory could conceivably have features (A)-(D). Klein (2013) suggests that semantic memory can, in principle, encode content of any kind. So, perhaps no content is necessary and sufficient for episodic memory. But the objection to the what-where-when view is not that it fails to articulate necessary and sufficient conditions for episodic memory. Episodic memory might well not be characterizable in terms of necessary and sufficient conditions; even if it is, it is no part of the project of detecting episodic memory in animals to provide them. The problem with the what-where-when view is rather that it characterizes episodic memory in terms of a content which many semantic memories have. There is nothing unusual about a semantic memory encoding what-where-when content. So, if animals do encode what-where-when content in memory, this does not obviously distinguish between the episodic and semantic memory hypotheses. The phenomenological view will be a companion in guilt with the what-where-when view only if the same is true for (A)-(D) – that is, only if these are features semantic memories commonly or easily have.

This does not seem to be the case: episodic and semantic memories appear standardly to have quite different contents. The ‘autobiographical interview’ protocol is a reliable and valid method for distinguishing episodic and semantic memory in humans. It distinguishes them by recording the number and type of details an individual can recall about a past event. Certain types of information are rated as episodic, others as semantic. Event related

---

10 At least, it shouldn’t be for the reasons given here. In places though, e.g. (Suddendorf & Corballis, 2007), it is presented this way.
information, details about time and place, and self-regarding information (about one’s position, for instance) are treated as episodic (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002). Correspondingly, individuals with severe episodic memory deficits are impaired against controls in this test – they recall fewer of these details, and those they do recall are less specific (Palombo, Alain, Söderlund, Khuu, & Levine, 2015). In tests of complex figure recall, episodically impaired individuals are also impaired on recall of visuo-spatial information (Palombo et al., 2015). So, it is reasonable to conclude that semantic memory does not standardly deal in this kind of information.

Having said that, individuals with severely deficient episodic memory can sometimes report on past events in some detail. Evidence regarding this phenomenon is scarce, but there are reasons for thinking that this is not overly troubling for the project of distinguishing episodic and semantic memory in non-humans. The individuals report that they achieve these feats of recall by ‘rehearsing’ the events to themselves or looking at photographs (Palombo et al., 2015). This suggests, first, that such ‘quasi-episodic’ semantic memories arise only where the subject makes a concerted effort to commit something to memory; they are unlikely to be formed by individuals who are not motivated, at the time of an event, to commit details about it to memory. Second, and more importantly, their reports suggest that forming and maintaining quasi-episodic semantic memories relies on mnemonic aids like language and photographs. Since these props are unavailable to nonhuman animals, it is doubtful that animals can form quasi-episodic semantic memories, even where humans can. So, although semantic memories conceivably can, and perhaps sometimes do, have features (A)-(D), the presence of these features in nonhuman memory should nevertheless be treated as evidence for nonhuman episodic memory.

How might we detect these features in non-verbal behaviour? Some existing tests already aim to detect some of these features. What-where-when studies seem at least partially to investigate (A) – recall of visuospatial features. Studies of sequence recall (e.g. Allen et al., 2014) investigate whether nonhuman memories exhibit (B). But neither of these approaches addresses (C) – they do not speak to whether the subject represents a past state of the world as past. In what-where-when studies, subjects might simply represent (say) that there are some worms at L, and that they have been there for a day, without recalling the event during which they were placed there. In this way, the scrub-jay’s memory may still represent the present state of the world, and not any previous event. Similarly, subjects might pass tests of sequence recall by representing ‘how the sequence goes’, rather than by recalling any past occasion on which
they were exposed to the sequence (Hoerl, 2008; see also McCormack, 2001). At issue in both cases is how to distinguish representation of a past event as such from mere sensitivity to one.

Unexpected question tasks may be the key to solving this problem. In these tasks, the subject is required to recall information which, at the time of the event, was ‘incidental’ – which the subject was not expecting to need (Fujita, Morisaki, Takaoka, Maeda, & Hori, 2012; Mercado, Murray, Uyeyama, Pack, & Herman, 1998; Singer & Zentall, 2007; Zentall, Clement, Bhatt, & Allen, 2001; Zentall, Singer, & Stagner, 2008; Zhou, Hohmann, & Crystal, 2012). The underlying thought is that, if information appears unimportant at the time of the event, it is unlikely to be encoded in semantic memory – but if the event is episodically remembered, the subject should be able to ‘revisit’ her experience to retrieve the relevant information. Presumably, the reason we might expect episodic memories to contain such incidental information is that episodic memories constitutively involve mental imagery, have apparent temporal structure and present events as self-involving – so, it is in their nature to record detailed visuo-spatial information, represent an event’s temporal features and have self-referential content.

Existing unexpected question tasks face two limitations. One is that they require the establishment of a form of non-verbal communication – so that the subject can be asked and can answer a question. The associated training could lead the subject to assign significance to the sort of information she is being asked to report on, and so to commit it to memory even where she is not expecting to be asked about it (Fujita et al., 2012). Another is that these tests typically ask subjects about an event which occurred only seconds ago – so it is not clear that they discriminate episodic from working memory (Zentall, 2013). A recent study with dogs used an alternative methodology to avoid these problems (Fujita et al., 2012). The retention interval was longer, and the subjects were not trained. Instead, the dogs explored an environment containing baited and unbaited containers, and were able to eat from some but not others. When they were unexpectedly returned to this environment, they seemed to remember which containers were baited, and where they had previously been. Yet it seems that this study redresses the issues with the unexpected question format at the cost of failing to address (C). Since the dogs were required to recall only which containers still contained food, their behaviour could have resulted from representing the present state of the world.

---

11 See also (Cheke & Clayton, 2013, 2015) for examples of verbal unexpected question tasks.
Nevertheless, a variant on the unexpected question format could be used to investigate (C). In an unexpected question task, the subject must recall information which was incidental at the time of the event. But to represent a past event as such is to represent it as superseded by, but causally and practically relevant to the present. To determine whether the subject represents the past event in this way, the information to be recalled in our unexpected question task should meet two further conditions. First, it should no longer reflect the state of the world – it should have been superseded, and the subject should be aware of this. This ensures that if the subject recalls the information, it is not part of her model of the present state of the world. Second, since the event, the information should have become relevant to the pursuit of the subject’s goals. This achieves two things. If, despite knowing that the information has been superseded, the subject recognises its present significance, this indicates that she represents the past event as past, in the sense outlined above. Moreover, by making the information unexpectedly relevant to the subject’s pursuit of her current goals, we can essentially ask her an ‘unexpected question’ after the fact, without training her in any system of communication.

Here is an example, to make things clearer. Suppose we are investigating whether scrub-jays have episodic memory. At t1, we present them with a target event to be remembered, in which they are allowed to cache food in an environment we have marked in some arbitrary way – with a patch on the wall, say. At t2, they learn that the patch is no longer present: so, the information about the patch has now been ‘superseded’. At t3, we present them with a learning event, in which they discover that if the patch was present at caching, their food will disappear from the cache. At t4, they return to the original environment. If they do not look for the food, this indicates that they recall the patch was present at t1. Since by t3, they already knew that the patch was absent, they cannot have used what they learned at t3 to update a present model of the world – since their present model of the world at t3 does not represent the mark being present. The information learned at t3 can inform subjects’ behaviour at t4 only on the assumption that subjects represent at t3 what was the case at t1, and appreciate its practical significance.

This sort of task could show that animals recall information about a past event as past, addressing (C), since success requires representing that something was the case, which is known not to be the case now. What about (A), (B) and (D)? These pick out the type of information episodic memories typically store about a past event – information about the event’s visuo-spatial features, its temporal structure and the subject’s involvement in the event. The sort of task I have proposed works by assigning arbitrary significance, after the fact, to some piece
information which the subject might be expected to recall if they can episodically remember – because the information is of a kind episodic memories store constitutively. In the above example, the information concerned the event’s visuo-spatial features – but it might equally have concerned the event’s temporal structure (the order in which things happened, say) or the subject’s involvement in the event (where she was located, perhaps). So, a task of the kind proposed here can be used to investigate whether subjects recall a past event as past, in terms of its visuo-spatial features, its temporal structure and her own involvement in the event. If this suggestion is workable, then each of the impure phenomenological features highlighted by the phenomenological account of episodic memory can be detected in non-verbal behaviour.

The proposal here resembles but differs from the methodology adopted in Clayton, Yu and Dickinson (2003), which found that scrub-jays can use information learned after caching to decide which cache to return to. The subjects in this study were not required to recall anything which was incidental at the time of encoding, nor did they need to recall anything which they knew to have been superseded in the interim. Experiment 1 in Crystal et al. (2013) comes very close to what I have in mind, however. In this experiment, rats were able to learn that food would replenish at a location only if they had originally discovered it there themselves, rather than having been placed there by an experimenter. That they were able to learn this rule suggests that, at the time of replenishment in the learning-phase, they recalled the circumstances under which they had originally found the food – perhaps in terms of the event’s perceptible features, or in terms of their own role in the event. At the time of encoding, these circumstances may well have been incidental. The methodology differs from that proposed here only in that this rule was learned before the occurrence of the events to be recalled in the test phase. At this point, having learned the replenishment rule, information about experimenter-placement would have ceased to be incidental: the rats would have clear motivation to remember the feature on which replenishment depends. Despite this difference, that the rats were able to learn the rule prior to the test phase provides a promising indication that their memories have some of the impure phenomenological features outlined here.12

12 [Acknowledgment removed.]
5. Conclusion

I have argued that the phenomenological features of episodic memory are *impure* phenomenological features – they have both subjective and objective aspects. So, to characterize episodic memory in terms of these phenomenological features is, in part, to highlight that it has certain objective features: it represents past events *as* past, recording detailed information about their visuo-spatial features, as well as their temporal structure and the subject’s own involvement. Consequently, this characterization does not make it impossible to test for episodic memory in nonhuman animals. Rather, it implies that we should investigate whether animals’ memories have these objective features – and if they do, that we should conclude that animals have episodic memory.

These objective features, I’ve argued, can be detected in nonverbal behaviour using methodologies similar to those already employed in episodic-like memory research. So, whilst my defence of the phenomenological view of episodic memory in one sense challenges the orthodoxy in comparative psychology, in another it is vindicatory. The rejection of phenomenological characterizations of episodic memory suggests a kind of pessimism about the power of behavioural tests to reveal the phenomenological features of nonhuman experience – indicating that they can warrant nothing stronger than agnosticism. In arguing that we can characterize episodic memory as ‘mental reliving’, and nevertheless detect it in animal behaviour, I have indirectly been making the case for a more optimistic view: that from behavioural tests we can rationally draw conclusions not only about what animals remember, but about what their remembering is like.

References


Klein, S. B. (2013). Making the case that episodic recollection is attributable to operations


