What is it like to be a baby?

Natural kinds and infant consciousness

Henry Taylor

j.h.taylor.1@bham.ac.uk

Draft of July 2022: please do not cite or distribute.

Word Count: 4980

Abstract

Studying consciousness in prelinguistic infants presents a challenge. We cannot ask them what they saw, and they cannot understand complex task instructions. This paper offers an optimistic methodology for studying infant consciousness, by drawing on philosophical work concerning natural kinds. I argue that this methodology is scientifically realistic. I also use it to interpret recent neuroscientific results concerning conscious perception in infants.

1. *Never work with children or animals.*

Studying consciousness in prelinguistic infants is crucial for understanding consciousness, and how it originates in the brain. But it presents a host of methodological challenges. In perceptual psychology, verbal report is the gold standard for establishing the presence of conscious experience. Clearly prelinguistic infants cannot verbally report their experience. Prelinguistic infants cannot understand complex task instructions, and thus cannot participate in the experiments that we routinely use to probe consciousness in adults. In short, there are severe methodological problems with applying our standard psychological and neuroscientific measures of consciousness to prelinguistic infants.

Faced with these problems, it is tempting to give up on studying consciousness in prelinguistic infants. Here, I take a more optimistic approach. In section 2, I explain why the dominant method for studying cognitive function in prelinguistic infants cannot be applied to study consciousness. I then (section 3) outline some work on natural kinds, concentrating on an inference style that I call co-projection. Section 4 argues that, by taking a natural kinds approach to consciousness, we can develop a methodology for understanding consciousness in infants. In section 5, I make the framework more concrete by relating it to recent work in neuroscience. I then (section 6) explain a particularly exciting aspect of the framework, which is that it promises to help us develop new measures for consciousness in infants. I conclude by briefly explaining how the approach deals with different ways that the empirical facts may pan out (section 7).

By 'consciousness' I mean phenomenal consciousness throughout, not access consciousness (Block 2007). I take consciousness to be a property of certain mental events (the ones that there is something it is like to undergo). Nothing in my argument hangs on this, and little changes if we take consciousness to be a property of the subjects themselves. By 'infants' I mean prelinguistic human infants. The topic of this paper is consciousness in infants. I will not address broader methodological problems with consciousness science that also apply to adult humans. For example, I will not discuss how consciousness should be empirically distinguished from cognitive access (Block 2007).

Treating consciousness as a natural kind has been suggested before (Shea and Bayne 2010, Shea 2012). However, this paper is the first attempt to use the approach to address problems with studying infants. Furthermore, I do not have space to answer general concerns about whether consciousness should be thought of in natural kind terms (see Bayne and Shea 2020 for replies to such criticisms). The purpose of this paper is to *apply* natural kinds to infants, not to defend the general approach of using natural kinds to understand consciousness.

2. *Problems with the current paradigm.*

One dominant paradigm in the study of infant cognition uses looking behaviours as its main behavioural measure. An event takes place in front of the infant, which would be surprising to an adult. The more the infant looks at the event, the greater their surprise is taken to be. From the fact that certain events surprise them, we can infer what their expectations were, and then draw conclusions about their cognitive make-up. For example, a 2.5-month-old infant might be shown a toy, which then disappears behind a screen (the screen is the 'occluder'). In one condition, there are two screens, with a gap between them. The toy disappears behind the left screen, then *does not* appear in the gap, and reappears walking out of the right screen. Infants look for increased time at this result. This is taken to imply that they found it surprising

(Aguiar and Baillargeon 1999). From this it is inferred that 2.5-month-old infants understand that an object should become visible when it passes between occluders.

This paradigm is primarily used to study infant *cognition*. Could it be used to study *consciousness*? The suggestion would be that we infer from the fact that an infant looked directly at some event, to the conclusion that they are conscious of that event. This inference is problematic. We cannot infer from the fact that a subject focusses their gaze on a particular event to the conclusion that they are conscious of that event. One reason this inference is problematic is the phenomenon of 'inattentional blindness', which shows that subjects can fail to notice large and obvious stimuli that fall centrally in their visual field (Mack and Rock 1998).

In response to these problems, it could be suggested that inattentional blindness only arises when there is attentional distraction, so *if* attentional distraction were eliminated, then looking behaviours would be a good indicator of consciousness in infants. This is problematic because there is no way to establish where an infant's attention is directed, and hence no way to establish whether infants are attentionally distracted. We cannot assume that the infant's attention is directed toward the same stimulus that they are looking directly at, since attention can freely roam to areas of the visual field other than where the eyes are focussed. Furthermore, many paradigms that establish attentional fixation in adults rely on complex task instructions, and thus cannot be applied to infants. Some measures of attentional fixation that have been applied to infants include changes in heart rate (Richards 2010). However, such changes measures do not tell you *where* the infant is attending. In any case, in adults, stimuli can be attended, and not be consciously perceived (Kentridge et al. 2008). So even if we could positively establish that the infant is looking directly at a stimulus, and attending that stimulus, it still would not be enough for us to conclude that she was conscious of it.

This is not a critique of current methods in developmental psychology that make use of looking behaviours. Such paradigms aim to measure cognitive complexity, rather than consciousness. Rather, the point is that these dominant ways of probing infants' cognitive architecture cannot be used to establish consciousness. We need something else.

3. *Natural kinds and co-projection.*

We will require some background on natural kinds. Essentialist accounts of natural kinds claim that all members of a natural kind share a set of necessary and sufficient intrinsic properties that are definitive of the kind. Essentialism is now out of favour for the natural kinds of biology and psychology (Taylor 2019). This paper will assume the more popular view, which is that natural kinds in biology and psychology are property clusters (Boyd 1989). According to this view, to be a member of kind K, an entity must instantiate a loose cluster of properties ($P_1 \ldots P_n$), which are underpinned by some underlying mechanism. For example, members the natural kind *polar bear* is instantiate properties such as white fur, a certain skeletal structure, a particular diet, etc.

The cluster is 'loose' in the sense that properties are not *necessary* for kind membership. Something can fail to instantiate one or more of ($P_1 \ldots P_n$), and still be members of the kind (e.g. a polar bear without any fur is still a polar bear). By allowing that no one property is necessary for kind membership, the view allows that natural kinds can be vague in their extensions. As certain members of a kind instantiate some but not all of the properties in the cluster, there will be a fuzzy area where it is not clear whether an entity is a member of that kind (Taylor 2019).

It's controversial how strongly to interpret the claim that must be underpinned by a mechanism (Slater 2015). For this paper, we can be quite liberal, and say only that ($P_1 \ldots P_n$) must cluster together for some reason that is at least in principle amenable to scientific investigation. In the polar bear case, the reasons that all these properties cluster together in the same organism will be to do with common ancestry and genetic inheritance. This is compatible with the claim that the mechanism could itself be another property, or that the properties in the cluster all mutually underpin one another. The intuition here is just that the properties should cluster together *for a reason.*

I do not have space for a full examination of the virtues of the property cluster view, and I cannot mount a defence of it from its critics. This paper will apply the view to infant consciousness. Even thinkers sceptical of the view will be interested to see how it can be applied to such a case.

The most relevant feature of cluster kinds, for our purposes, is that they support what I call *co-projection inferences*. Co-projection is where you infer from the fact that an entity instantiates some property

P_1 or set of properties P_1, P_2, P_3 to the conclusion that it probably also instantiates some other property in the cluster P_n. For example, if we know that some entity has a density of 19.3gcm^{-3} and that it has a melting point of 1337K, we can infer that it is probably an instance of gold, and hence that it will also have the property of having atomic number 79 (Khalidi 2018). The cluster kind view can easily explain how natural kinds ensure the reliability of co-projective inferences. Suppose that the cluster includes properties ($P_1 \ldots P_n$), and that they are all supported by some common mechanism M. The presence of *some* of the properties in ($P_1 \ldots P_n$) raises the chances of other properties being present, because all of the properties in the set ($P_1 \ldots P_n$) are supported by the same mechanism M. That is, the presence of P_1, P_2, P_3 raises the chances of M being present, because M underpins P_1, P_2, P_3, which in turn raises the chances of P_n being present, because M itself underpins P_n. I summarise this inference pattern in the following box:

Co-projection inferences:
The presence of a large number of properties distinctive of a cluster kind reliably indicates the presence of other properties in the cluster.

The word 'large' is intended to capture the idea that the more of properties in the cluster ($P_1 \ldots P_n$) are present, the higher the chances that M is also present, and thereby that the other properties in the cluster are present. 'Large' doesn't mean that there is some precise threshold number of properties above which inferring other properties in the cluster becomes reliable.

4. *How to study consciousness in human infants.*

How can this help with infants? In brief, we take consciousness to be one of the properties in the cluster of properties that is definitive of a natural kind. We use experimental studies on adults to identify the properties that co-occur with consciousness. This serves as a basis for inferring that these are the *other* properties in the cluster definitive of a natural kind. We would then have a picture of the cluster of properties, of which consciousness is one. The co-projective inference pattern explained above gives us good reason to infer from the presence of the *other* properties to the presence of consciousness (because they are all members of the same cluster). If we can establish the presence of those other properties in infants, then we would have a basis to infer the presence of consciousness, even in the absence of verbal report.

In more detail, the approach proceeds in three steps.

> **Step one:** use experiments on adults to identify the properties that cluster together with consciousness of some stimulus.

Essentially, in the first step, we are establishing which properties cluster together with consciousness. This will provide us with a list of properties P_1-P_3 which are in the same property cluster as the consciousness property C.

> **Step two:** identify which of these properties can be tested for in infants.

Some properties that correlate with consciousness in adults may be properties that we cannot test for in infants. For example, some properties that reliably cluster together with consciousness in adults may only be testable given complex task instructions that infants cannot understand. However, some properties in the cluster will be suitable for testing in infants (section 5 will give examples).

> **Step three:** test for the presence of those properties in infants, and infer from the presence of them to the conclusion that the infant is conscious of a particular stimulus.

Once we have identified properties P_1-P_3 (the ones that co-occur with consciousness of a particular stimulus in adult humans) we can use the presence of those properties in infants to infer the presence of C (consciousness of that stimulus). The reliability of this inference is grounded in the fact that natural kinds support reliable co-projection inferences.

5. *A concrete case.*

This is all very abstract. How would it look in practice? Step one will require widely accepted paradigms that can manipulate the presence and absence of conscious perception of a stimulus. Meta-contrast masking is an example (Kentridge et al. 2008). In this technique, a prime is presented, which is a coloured disc. This prime then disappears, and is followed after a brief interval with the presentation of an annulus (a coloured ring). The inner contours of the annulus (ring) coincide with the outer contours of the prime (disc). The result is that (if the temporal interval between the prime and the ring is just right) then the signals associated with the annulus will 'override' the signals from the prime in the visual cortex, and the prime will not be consciously perceived. Only the annulus will be consciously perceived. This is the 'consciousness absent' condition, because the prime is *not* consciously perceived. In the 'consciousness present' condition, the temporal interval between the prime and annulus is extended, and the effect disappears, meaning that the prime *and* the annulus are consciously perceived.

What this provides is a case where in some conditions the subject is conscious of the prime, and in others they are not. In this paradigm (deployed on *adults*), the presence of the prime is established by subjective report. Conversely, the absence of consciousness of the prime is established when the subject cannot report the prime. One of the virtues of meta-contrast masking is that it presents a (relatively!) uncontroversial way to manipulate the presence and absence of conscious perception of a stimulus. With this in hand, we can then look for the properties that correlate with conscious perception of the stimulus. That is, find the properties that are present when the stimulus is consciously perceived and absent when it is not. In terms of natural kind theory, we are identifying P_1-P_3: the properties that co-occur in the same cluster as the consciousness property C.

For this to be useful for infants, we cannot identify any old properties that co-occur with consciousness. As step two indicates, we need the ones that co-occur with consciousness *and* which can be safely tested for in infants. A good place to look would be neuroscientific properties. Specifically, we can use EEG and fMRI on young infants (Wu et al. 2019, Ellis et al. 2020). We should therefore look for the neural properties that meet two criteria: they correlate with consciousness, and they are detectable using EEG and fMRI. To do this, we do EEG and fMRI scans on *adult* subjects undergoing meta-contrast masking. We look at the neural properties that occur in the consciousness-present condition of the meta-contrast masking experiment, and which are absent in the consciousness-absent condition. This gives us a list of neural properties P_1-P_3, that co-occur with consciousness. With this list in hand, only now do infants enter the picture. We expose infants to stimuli, while they undergo fMRI and EEG. If we find the same properties in the infant, then (based on the co-projective nature of natural kinds) we are justified in inferring that the infant consciously perceived those stimuli.

Meta-contrast masking will not be the only psychological paradigm we use. We will need convergent evidence from a wide range of paradigms that represent (relatively!) uncontroversial ways of manipulating the presence and absence of consciousness. Another candidate technique is binocular rivalry (Fang and

He 2005), though I do not have space to discuss it. Together, all of these paradigms will allow us to distil the other properties in the cluster of which consciousness is also a member.

Kouider et al. (2013) represent a good example of an approach that fits the natural kinds framework. They identified a set of event related potentials (ERPs) that are correlated with conscious perception in adults (as established by a range of psychological paradigms). ERPs are electrical neural events, measurable (over enough trials) using EEG. The ERPs in question were associated with activation of anterior cingulate and fronto-parietal cortices in adult humans. Kouider et al. (2013) exposed infants to pictures of faces, and discovered the presence of similar ERPs across these areas of the cortex in infants. They did, however, find that infant brains took longer to generate the relevant ERPs (900ms in 5-month-old infants, compared to 300ms in adults).

The authors themselves deny that this would be a proof of conscious experience in infants, suggesting scepticism about whether neuroscientific measures could ever provide such proof (2013, 380). The natural kinds approach allows us to provide justification for the inference from the ERP data to conscious experience. What the results show is the presence of a property that we have good reason to think is present in the same natural kind cluster as consciousness. This fact can then be combined with our knowledge of the reliability of co-projection (which is a general feature of natural kinds, see section 3) to justify the inference to the claim that prelinguistic infants are conscious of the relevant stimulus.

In other words, the inference is as follows:

1) These ERPs are probably properties in the same natural kind cluster as consciousness of the stimulus.
2) The presence of some properties in a natural kind cluster raise the chances of the other properties being present.
3) (Therefore) The presence of these ERPs in infants raises the chance that the infant is conscious of the stimulus.

The justification for (1) comes from the fact that these ERPs are reliably correlated with consciousness in adults. The justification for (2) stems from our knowledge of the reliability of co-projective inferences in natural kinds generally. (2) is the distinctive epistemic contribution of the theory of natural kinds. On this picture, our general understanding of the reliability of co-projection (that is a general feature of natural kinds) is what provides justification for the second premise in the above reasoning. That is, we are not simply taking a property that correlates with consciousness in adults, and then assuming that this property will also correlate with consciousness in infants. Rather, natural kinds theory is specifically providing our justification for thinking that the properties correlated with consciousness in adults will also be correlated with consciousness in infants. The justification for this inference is that the properties probabilistically group together into a cluster kind. As such, they will support a co-projective inference, which provides the justification for us to infer from some members of the cluster to others.[1]

These data only provide evidence of *some* properties associated with consciousness in adults (ERPs indicative of activation in fronto-parietal and anterior cingulate cortices). The probability that infants are conscious of the stimuli in question is raised by finding this distinctive pattern of ERP activity. However, it would be preferable if we could establish the presence of more of the properties in the cluster distinctive of consciousness, as that would raise the chances of the consciousness property being present. We can do this by identifying further neural properties correlated with consciousness (identifiable using fMRI and EEG) and again looking for those in infants. The more properties we find, the higher our credence that the infants are conscious of the stimulus in question.

Of course, this framework can only ever give us probabilistic reasons to posit the presence of consciousness. It cannot give us certainty. That's no problem, since it is unreasonable to ask for certainty in most areas of science.

6. New measures.

Applying natural kinds, understood as cluster kinds, to infant consciousness also supplies another exciting possibility, which is that it can generate new ways of testing for consciousness in infants.[2] Suppose our

[1] The framework would operate in a similar way if it turns out that consciousness is not one of the properties in the property cluster, but is the mechanism that underpins the property cluster itself (cf. Shea 2012). In such a case, the presence of properties in the cluster would still reliably indicate the presence of consciousness, and hence the presence of these properties would still allow us to infer the presence of consciousness. The rest of the framework would still apply.

[2] Compare Shea's discussion of natural kinds and measures of consciousness in adults (2012).

cluster of properties includes P_1-P_3 and the consciousness property C, such that the presence of P_1-P_3 gives us good reason to infer the presence of C. Now, suppose that we observe the presence of P_1-P_3 in infants, and thereby have good reason to infer the presence of C. Suppose also that we notice another property that is correlated reliably with this cluster in infants. This will give us (defeasible!) reason to think that this other property P_4 correlates with consciousness of the stimulus in question. We can then expand our set of measures of consciousness to include P_4. That is, we could take P_4 to be an indicator of consciousness. We would have used the natural kinds framework to establish a new measure of consciousness in infants.

Such a property might be another neural property, but it could in principle be any kind of property, including a behavioural one. Consider the following hypothetical example. Suppose we establish that a certain pattern of event related potentials (ERPs) are correlated with consciousness in adults. We then measure infant brain activity, and discover that the relevant ERPs are observed when certain stimuli are presented to them. We thereby conclude that they are conscious of those stimuli. Suppose also that we find that this particular pattern of ERP activity correlates with a specific behavioural trait. For example, suppose that in 3-month-old infants, it correlates with the infant fixating the stimulus for longer than 400ms. We can then incorporate this property into the cluster of properties that is definitive of the kind. With this in hand, we now have a behavioural property that correlates with consciousness in infants. We could conclude that (*certis paribus)* if the infant fixates the stimulus for longer than 400ms, this raises the likelihood that they are conscious of the stimulus.

7. *More complicated cases.*

In closing, I will discuss three more complicated empirical possibilities, to demonstrate that the framework has the resources to handle them.

7.1: What if infants don't have any of the properties?

The first case is if we identify the cluster of properties that correlate with consciousness in adults, then find that infants have few, or none of them. Here, the answer that the framework would give us is that we have no evidence that the infant is conscious. In such a case, simplicity dictates that (in the absence of any other positive reason to think they are conscious) we should conclude that they are not.

This result is overwhelmingly unlikely. We already have excellent evidence that the neural systems underpinning conscious perception in infants are continuous with those of adults (Kouider et al. 2013). Unlikely as this result is, it is a strong point in favour of the framework that it can in principle deliver the conclusion that infants are not conscious. After all, there would be something seriously wrong with an empirical approach that decided in advance that its subjects were conscious!

7.2: What if infants only have some of the properties?

More empirically likely is that there are cases where the infant possesses some but not all of the properties in the cluster that also contains consciousness. For example, let's say that P_1, P_2 and P_3 are the other properties in the same cluster as consciousness, and we discover (through neuroscientific techniques) that infants in a certain age range have properties P_1 and P_2, but not P_3. What then?

There would be several hypotheses available to us, and several ways to test for them. We would first try to establish whether P_3 really should be in the cluster in the first place. We would look for cases (in adults) where P_1 and P_2 are present, but P_3 is not. If we can find such a case, we can determine whether such subjects pass normal experimental criteria for consciousness of some stimulus (verbal report, for example). If they do, we can conclude that P_3 does not correlate very well with consciousness after all, and hence drop it from our cluster of properties indicative of consciousness. In such a case, the fact that infants do not instantiate property P_3 would not matter very much, because we would have been wrong to include it in the cluster that co-occurs with consciousness in the first place.

There may be more difficult cases where this kind of dissociation is not possible in adults, and so we are still confronted with the question of what to say about the infant that has P_1 and P_2, but not P_3. In such a case, we should simply accept that P_1 and P_2 raise the chances of the presence of consciousness in infants, but that the probability is simply lower than it would be if we had all three properties. In such a case, the cluster of which consciousness is a part would not be entirely present, so our basis for inferring the presence of consciousness is weakened. Such a result should simply be accepted.

This is a feature of the framework, not a bug. It is already known that there are cases within psychology where some properties distinctive of a cluster kind are present, and others are not. It is established that such results can carry important upshots for consciousness science (e.g. Taylor (2019)). Such results should just be embraced as part of the messy nature of the world.

7.3: What if consciousness is not associated with one natural kind?

I have assumed throughout that the consciousness property would be one property in the cluster definitive of a cluster kind. However, it might be that consciousness is actually associated with several distinct natural kinds (cf. Irvine 2013). It might be that, in adults, consciousness is actually present whenever *either* P_1-P_3 *or* P_4-P_6 are present. Suppose, for example, that visual consciousness is part of a cluster associated with P_1-P_3, and auditory consciousness is associated with a completely separate cluster of properties P_4-P_6. In that case, consciousness would be a disjunctive kind, in the sense that it would be associated with more than one natural kind.

Regardless of the empirical likelihood of this case, the framework has the resources to address it. We would simply conclude that the presence of P_1-P_3 in prelinguistic infants is indicative of visual consciousness of some stimulus, whilst P_4-P_6 is indicative of auditory consciousness of a stimulus. A more complicated result, perhaps, but obviously not one that presents a challenge to the framework.

8. *Conclusion.*

I propose using the machinery of cluster kinds to develop a framework for studying prelinguistic infant consciousness. Our knowledge of the ways that natural kinds reliably support co-projection supplies us with the justification to infer from the presence of other properties to the presence of consciousness. The framework also promises to expand our list of measures of consciousness, and thereby provide new ways of establishing consciousness in infants. The framework is empirically realistic and sidesteps the notorious methodological problems with studying infant consciousness.

References

Aguiar, Andréa. and Baillargeon, Renée. 1999. 9-month old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology*. 39: 116-157.

Bayne, Tim. and Shea, Nick. 2020. Consciousness, concepts and natural kinds. *Philosophical Topics.* 48: 65-83.

Block, Ned. 2007. Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioural and Brain Sciences.* 30: 481-548.

Boyd, Richard. 1989. What realism implies and what it does not. *Dialectica.* 43: 5-29.

Ellis, Chase, Skalaban, Lena, Yates, Tristan, Bejjanki, Vikranth, Códova, Natalia, Turk-Browne, Nicholas. 2020. Re-imagining fMRI for awake behaving infants. *Nature Communications.* 11:4523

Fang, Fang. and He, Shang. 2005. Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience.* 8: 1380-1385

Irvine, Elizabeth. 2013. *Consciousness as a Scientific Concept: a philosophy of science perspective.* Dorechdt: Springer.

Kentridge, Robert, Nijober, Tanja, and Heywood, Charles. 2008. Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia.* 46: 831-69.

Khalidi, Muhammed. 2018. Natural kinds as nodes in causal networks. *Synthese.* 195: 1379-1396.

Kouider, Sid, Stahlhut, Cartsen, Gelskov, Sophie, Barbosa, Leonardo, Dutat, Micahel, de Dardelle, Vincent, Christophe, Anne, Dehaene, Stanislas, Deheane-Lambertz, Ghislaine. 2013. A neural marker of perceptual consciousness in infants. *Science.* 340: 376-380.

Mack, Arien, and Rock, Irwin. 1998. *Inattentional blindness.* Cambridge, MA: MIT Press.

Richards, John. 2010. The development of attention to simple and complex visual stimuli in infants: behavioural and psychophysiological measures. *Developmental Review.* 30: 203-219.

Shea, Nick 2012. Methodological encounters with the phenomenal kind. *Philosophy and Phenomenological Research.* 84: 307-344.

Shea, Nick and Bayne, Tim. 2010. The vegetative state and the science of consciousness. *The British Journal for the Philosophy of Science.* 61: 459-484.

Slater, Matthew. 2015. Natural kindness. *The British Journal for the Philosophy of Science.* 66: 375-411.

Taylor, Henry. 2019. Fuzziness in the mind: can perception be unconscious? *Philosophy and Phenomenological Research.* 101: 383-398.

Wu, Joyce. et al. 2019. Scalp EEG spike predict impending epilepsy in TSC infants: a longitudinal observational study. *Epilepsia.* 60: 2428-2436.