PhiMiSci
Philosophy and the Mind Sciences

Ontogenetic emergence as a criterion for theories of consciousness
Comparing GNW, SOMA, and REFCON

Asger Kirkeby-Hinrup[a] (akh@cfin.au.dk)
Morten Overgaard[b] (morten.storm.overgaard@cfin.au.dk)

Abstract
In recent years increasing attention has been given to systematic comparison of theories of consciousness. Laudable practical projects have emerged in this regard, such as adversarial collaboration and the development of databases lending themselves to comparisons of empirical support for theories. In addition to the practical advances, theoretical advances have been made, such as a list of issues a theory of consciousness must address. We propose adding the issue of the ontogenetic emergence (O-emergence) of consciousness to the list of issues we use to evaluate theories of consciousness. O-emergence concerns how and when consciousness emerges ontogenetically in human beings. The underlying assumption is that there exists a point in the development of a human individual before which that individual is not and cannot be conscious. This assumption, in turn, depends on a widely shared assumption of cognitive neuroscience, which is that consciousness somehow depends on — or derives from — brain activity. In this paper, we lay out the O-emergence criterion and investigate whether it can be accounted for within the Global Neuronal Workspace theory, the Self-Organizing Meta-representational Account, and the Reorganization of Elementary Functions framework.

Keywords
Consciousness · Criteria for theories · Emergence · Ontogenetics

1 Introduction

When it comes to consciousness, some questions are notoriously difficult to answer. *How* does consciousness arise from brain activity? *Why* did consciousness emerge at all? *What*, if anything, does consciousness enable us to do, that we could not do without it? The first two of these questions correspond roughly to the so-called *hard problem* of consciousness (Chalmers, 1995). When viewed in light of metaphysics, the last question essentially concerns the causal efficacy

[a] Lund University, Sweden.
[b] Aarhus University, Denmark.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902
 https://philosophymindscience.org ISSN: 2699-0369

of consciousness and is sometimes framed as the *function* of consciousness (e.g. Rosenthal, 2008, 2012). While a consensus on these three questions has so far eluded the scientific community, we are nevertheless making progress in a closely related domain: assessing the candidate explanations of consciousness. In recent years, various efforts have been made to develop ways of comparing candidate explanations of consciousness (e.g. Doerig et al., 2020; Kirkeby-Hinrup & Fazekas, 2021), as well as considering the kinds of evidence necessary to moving forward debates (Fink, 2016; Kirkeby-Hinrup, 2024; Overgaard & Kirkeby-Hinrup, 2021a). While these efforts differ in scope and methodology, they share certain features: a motivation to deliver structure and focus to the debates, meta-considerations of evidence, and pointing to ways to proceed from here. One initiative to this effect, proposed by Adrien Doerig and colleagues (Doerig et al., 2020), advocates a range of criteria an empirical theory of consciousness must satisfy to be viable. Many of the criteria proposed by Doerig and colleagues pose challenges only to particular theories. Nevertheless, collecting these criteria into a comprehensive list is a commendable and (to our minds) necessary effort to establish an overview — and commence a discussion — of what we expect of a theory of consciousness.

We propose an additional criterion against which to compare theories of consciousness. The criterion in question concerns the ontogenetic development of consciousness (*O-emergence* for brevity). This criterion has been catapulted into the spotlight by the recent proposal of a theory (The Self-Organizing Meta-representational Account, or "SOMA") by Axel Cleeremans and colleagues (Cleeremans et al., 2020), which takes its point of departure in hypotheses about the stages involved in the acquisition of consciousness in early childhood. By offering an account of the acquisition of consciousness in ontogenetic development, SOMA has implicitly (but not explicitly) challenged proponents of competing theories of consciousness to show how their preferred theory can account for this.

In the couple of years after the publication of SOMA (but seemingly independent of SOMA) the prospects of consciousness in early life has received increased attention. Consequently, several different approaches and a variety of considerations have been offered. These range from how to measure perinatal experience (Frohlich et al., 2023), the role of co-embodiment in utero in shaping and scaffolding perceptual experience (Ciaunica, Constant, et al., 2021; Ciaunica, Safron, et al., 2021), as well as considerations about what extant theories of consciousness predict with respect to fetal consciousness similar to our endeavor here (Mudrik et al., 2023; Passos-Ferreira, forthcoming).

To preempt misunderstandings, we need to clarify that our usage of *emergence* pertains to how consciousness emerges in the *ontogenetic* development of human beings, i.e. how it emerges in the development of individuals. Thus, our use of emergence should not be conflated with another widespread use of the word,

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

which pertains to the *phylogenetic* development of consciousness in the human *species* (Cabanac et al., 2009) [1].

Returning to O-emergence, the underlying assumption is that there exists a point in the development of a human individual before which that individual is not conscious. This assumption, in turn, depends on a widely shared assumption in cognitive neuroscience, which is that consciousness (at least partly) depends on brain activity. In other words, most people involved in interdisciplinary consciousness studies think that the brain is *necessary* for consciousness[2], call this the *neural assumption*. If one accepts the neural assumption, then the existence of a *before* and *after* the appearance of consciousness follows because in early prenatal development, e.g. in the germinal stage (where the entity in question is a zygote consisting of only a few cells) the brain plainly has not physically formed yet. While the germinal stage serves to establish a clear cutoff before which (according to the ground assumption) there can be no consciousness because there is no brain, few believe that the consciousness emerges immediately with the start of brain development in the embryonic stage. Rather, the majority (but in no way universal) consensus in the field appears to be that certain features (e.g. processes, areas, or connections) of the brain are necessary for consciousness, and that these features mostly are not present until in later stages of prenatal development, or possibly early years of childhood. Thus, there is significant theoretical wiggle room with respect to hypotheses about O-emergence.

Answering the question of *when* consciousness emerges in human individuals invites the further question of *how* consciousness emerges at that point in the ontogenetic development. In other words, if a theory suggests that consciousness emerges around the age 85, we next want to know what normally happens in individuals (or their brains) around the age of 85, that makes plausible the claim that this is when consciousness emerges. Thus, addressing O-emergence consists in answering the *when* and *how* of the emergence of consciousness in ontogenetic development.

[1] Importantly, we do not claim that the evolutionary origins (phylogeny) of human consciousness is of no relevance to questions of O-emergence and vice versa. However, the assessment of theories based on O-emergence we propose can be done in a vacuum, as it were.

[2] It is worth noting that the three theories discussed here, as well the 'prominent ones' mentioned in clarification four below, all belong to what is sometimes called 'neuro-centric' theories of consciousness. Roughly speaking, neuro-centric means that a theory holds that a (functioning neuro-typical) brain is both necessary *and sufficient* for consciousness. We acknowledge that this is not only one view in the field, and that a range of theories are on offer that do not think the brain is *sufficient* for consciousness, but rather hold that there are non-neural necessary conditions (e.g. embodiment) that must obtain as well (see, e.g., Ciaunica, Safron, et al., 2021; Delafield-Butt & Gangopadhyay, 2013; Delafield-Butt & Trevarthen, 2022). Importantly, the O-emergence challenge pertains also to these theories. However, qua their nature, the explanation they provide will not exclusively be cached in neural terms. For this reason, it is likely that the questions used to evaluate O-emergence in relation to non-neuro-centric theories may require different formulations. We thank an anonymous reviewer for prompting this clarification.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

In section three, we propose using three questions to analyze and assess how a theory accounts for O-emergence. After this, we proceed to evaluate SOMA, the Global Neuronal Workspace theory (GNW) (Dehaene & Naccache, 2001), and the REFCON framework (Mogensen & Overgaard, 2018a) using the three questions. However, to cast this paper and the assessment of the three theories in the right light, in the next section, we start by laying out some clarifications.

2 Clarifications

Firstly, some theories might argue that O-emergence is a graded process, where consciousness slowly emerges over time. In those cases, how do we pinpoint the emergence of consciousness? Now, if one accepts the following premises 1) the brain is necessary for consciousness, 2) at the germinal stage there is no brain, and therefore can be no consciousness, 3) neurotypical adult humans are conscious, and 4) that time (at the level of description of individuals) is unidirectional, one is committed to there being a before and an after, i.e. that there is a point in time of ontogenetic emergence. In other words, one is committed to the graded process having a beginning. The place on the timeline to affix ontogenetic emergence would then be at exactly the point where the graded process begins, i.e. at the first point in time at which there is *something it is like for* (Nagel, 1974) the individual in question. Or put more colloquially, the first point in time where a hypothetical phenomenal consciousness meter reads above zero. In fact, one way to understand the definition of ontogenetic emergence is exactly this, namely the *first* transition from a point in time where the individual *has no* (phenomenally) consciousness experiences to a point in time where the individual *has* (phenomenally) conscious experiences. Or using the subjectively accessible marker of consciousness proposed by Nagel, the *first* transition from a point in time where there is *nothing it is like for* an individual to a point in time where there is *something it is like for* an individual.

In light of this, it is hard to see a non question begging argument for why we should *not* place O-emergence at this point in time. The reasoning goes like this: claiming that O-emergence should not be located at the beginning of the graded process entails that one holds that the early stages of the graded process should be excluded for one reason or another. For instance, one might claim that early stages of the graded process are a special kind of proto-conscious states that do not entail that there is something it is like to be the individual and that 'real' phenomenally conscious states only occur at a later point in the graded process, and this is reason to not pinpoint O-emergence to the beginning of the graded process. The reply to this argument is straightforward and consists simply in pointing out that the *first* transition from a point in time where the individual is in no conscious states to a point in time where the individual is in a proto-conscious state, according to the definition deployed in the argument, does not coincide with the transition between the *first* point in time where the individual is in no phenomenally conscious

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

state to a point in time where the individual is in a phenomenally conscious state. In other words, the important thing is that if there is nothing it is like to be in the proto-conscious states, then for the purposes of determining the point of O-emergence what counts is exactly that there is nothing it is like to be in them, i.e., that they are *not* phenomenally conscious states. The *first* transition from a point in time where the individual is in no conscious states to a point in time where the individual is in a proto-conscious state is irrelevant to determining O-emergence. In relation to O-emergence it is of no consequence whether at the temporally prior point the individual is in no conscious states at all, or if the individual has some proto-conscious states. That the proto-conscious states may be causally relevant to O-emergence also is of no consequence here because causal relevance is not a relevant property for determining O-emergence. The relevant property is whether the individual is phenomenally conscious, i.e. whether there is something it is like to be the individual, or not. In sum, arguing that there is a graded process, and that O-emergence occurs somewhere *after* the beginning of that process is tantamount to simply moving the starting point of the *relevant* graded process to a later point in time. Anything else seemingly entails working with another definition of O-emergence than the one committed to when accepting the four premises. Thus, invoking a graded process does not warrant dismissing the O-emergence criterion.

The second clarification concerns cases where one posits intermittent timespans of consciousness separated by periods of unconsciousness in the early stages of ontogenetic development. To illustrate this idea, suppose the fetus has brief bursts of conscious experience in the womb between which it returns to being a fully unconscious individual for extended periods of time (e.g. a week or a month). The crux is the supposition that these brief bursts of conscious experiences are not equivalent to the pervasive phenomenon that consciousness is for adult human beings but are instead stages in — or precursors to — the ontogenetic development of "persistent consciousness", as it were. The question now is: in cases where intermittent bursts of conscious experience precede a fully-fledged conscious mental life, where do we place the point of O-emergence? The answer to this question follows the same lines as the one given above when the question concerned a graded process. Because O-emergence is defined as the *first* transition between a point in time before which there has been nothing it is like to be that individual, and a point in time where — for the first time — there is something it is like to be that individual, O-emergence belongs at the starting point of the first burst of conscious experience. The reasoning for this mirrors the one given above. To reiterate, the proponent of intermittent bursts of consciousness needs to deliver a separate argument for *why* O-emergence should not coincide with the first burst of consciousness. However, such a separate argument will *either* beg the question against the definition of O-emergence or argue that there is nothing it is like for an individual to undergo these bursts of consciousness, which would render these bursts irrelevant to O-emergence.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

The third clarification is that the analysis we here offer of O-emergence takes for granted the conceptual framework of the theory under consideration. This means we are not (at present time) out to criticize theories or leverage objections to their way of conceiving of neither consciousness nor the processes each theory suggests is underpinning it. Rather, we will investigate theories on their own terms. In other words, what we aim to do here concerns *mapping* and *projecting*. In cases where a theory has provided answers (or indications) of any facet of O-emergence, we will faithfully map these while taking for granted the concepts the theory deploys. Where there are lacunas in the answers, we attempt to project the likely stance of a theory, again staying faithful and charitable to the conceptual framework of the theory. This practice should not be taken as an endorsement of the theories. In fact each author is ambivalent toward at least one theory. Neither should it be taken as an indication that we do not think there are good objections to some theories' conceptual frameworks, conceptions of the explanandum (consciousness), or their respective accounts of O-emergence. The critical point is that before one can rightfully criticize a theory, it is imperative to understand the actual claims of the theory, as they are espoused by its proponents. Thus, the present paper is about setting the foundation on which to have a debate (that may involve criticism) about O-emergence. The aim is understanding what theories claim (or are likely to claim) and examining upshots of these claims without problematizing or objecting to them.

The fourth clarification concerns the selection of theories treated here. Importantly, every candidate explanation of consciousness needs to be able to account for O-emergence, and therefore should be subjected to the kind of evaluation we here apply to SOMA, GNW and REFCON[3]. Therefore, while we only discuss three theories in this text, we want to underscore that the remaining prominent hypotheses of consciousness, such as — among others — the Local Recurrency Theory, Higher-Order Thought Theory, and Information Integration Theory (see e.g. Lamme, 2004; Lau & Rosenthal, 2011; Tononi, 2005 respectively) should undergo a similar process of mapping and projecting[4].

[3] Recently, Ciaunica et al. (2021) have made important progress in this regard. Furthermore, it is worth noting that the domain of non-neuro-centric theories in general seem to be ahead of its neuro-centric counterpart with respect to discussing O-emergence.

[4] Two objections have been leveraged to the selection of theories we consider here. The first objection holds that we should focus on the most prominent theories, and that the SOMA and REF framework do not fall into this category. The second objection is that the three theories treated here are too dissimilar and treating them as the same kind of theory is counterproductive. Let us address each of these objections in turn.

Our reply to the first objection is that it perpetuates a kind of socio-scientific celebrity bias. To elaborate, one reason (among others) the prominent theories — e.g. Integrated information theory (Albantakis, 2020), higher-order thought theory (Brown et al., 2019), recurrent processing theory (Block, 2007) — are prominent is that many researchers are interested in them and publish about them (in fact, this is roughly what prominent means). We therefore consider it likely that these theories will be discussed in relation to O-emergence even if they are not included here (something that is less likely to be the case with SOMA and REFCON). This assumption

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

The fifth clarification concerns the core of the subject matter, namely consciousness. To avoid misunderstandings, it is necessary to be clear about how we conceive of the phenomenon in the present context. The initial answer is simple: the conception of consciousness that we operate with in treating a given theory is inherited from the theory. Doing this is part of our endeavor to charitably assess the theories *granting their conceptual frameworks*. Critically, one may now point out that the three theories considered here do not conceive of consciousness in the same way, so we are comparing apples with oranges. Put differently, one might object that evaluating their stances on O-emergence would be inherently flawed

is supported by the fact that it is significantly easier to publish discussions of the prominent theories for the simple reason that they are prominent, and therefore assumed to have a wider readership and therefore be more attractive to journals. However, the upshot of this is that it is comparatively more difficult to publish about the non-prominent theories (the fact that this objection is discussed is partly proof of this). Thus, prominence seems to be self-perpetuating. Relatedly, the fact that the prominent theories have larger bodies of empirical support is a natural consequence of the additional attention these theories have enjoyed, under the assumption that increased attention means more research is being done (i.e. attention correlates with time and money). More importantly, given that no-one yet knows what the case with human consciousness (and its relation to the brain) is, one should be careful not to conflate popularity with truth or plausibility. In sum, we are confident that our selection of theories in this text is of no grave concern, since — given their popularity — the prominent theories are likely to be discussed in relation to O-emergence eventually. The main aim of this paper exactly is to provide the foundation for such future discussion.

Turning to the second objection that the three theories to be considered here are too dissimilar to be subjected to the same criteria, we submit that this is a feature, not a bug. We readily concede that GNW, SOMA and REFCON are indeed dissimilar in several and significant ways. GNW as the reader likely will know, is a widely discussed theory subjected to several decades of empirical work, and with a theoretical foundation stretching back to the work of Barnard Baars in the late eighties. SOMA by contrast is one of the newest additions in the field of interdisciplinary consciousness studies, with a comparatively miniscule body of work. Furthermore, some may see the SOMA as more of a programmatic approach than an actual theory. While we disagree with this characterization, it is of no consequence here for the reasons given below. Finally, REFCON is a theory that grew out of an attempt to solve an entirely different problem and its body of work pertaining to consciousness, while larger than that of SOMA, is still limited. Finally, GNW has a few papers with neurological considerations pertinent to O-emergence, SOMA presents a theoretical framework mapping hypothesized stages in the acquisition of consciousness, and REFCON has neither. So, having conceded that these theories are indeed different both in scope, focus, approach, and size of their bodies of work, what merits subjecting them to assessment using a single set of criteria? The answer is twofold. Firstly, despite their differences, all three theories are about consciousness. Secondly, their dissimilarities are exactly the reason they were chosen. The different foci and level of development (both in general and with respect to O-emergence) of GNW, SOMA and REFCON will serve to illustrate how assessing O-emergence can shed light on overlooked implications, offer new perspectives, and point to future avenues of research. Since we are attempting to set the framework for discussions of O-emergence, selecting significantly different theories here doubles as a proof of concept for an O-emergence research program by showing it is applicable to many different kinds of theories of consciousness. Furthermore, its applicability to theories with different scopes or foci is useful, not only for assessing and comparing theories across the field, but additionally for raising new questions and illuminating strengths or shortcomings in relation to each theory.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

exactly because the phenomenon they each hold to O-emerge is not the same (this objection is a variation on the ones discussed above concerning dissimilarities between the theories). This is not an unreasonable objection, but it misconceives what is the focus of our treatment here. To elaborate, most theories in the field agree that there is *something it is like* to be conscious, and this is the focus when considering O-emergence. Indeed, referring to Nagel's way of capturing the phenomenon is prevalent in the literature (for just some recent examples see e.g. Blum & Blum, 2022; Frohlich et al., 2021; Northoff & Lamme, 2020; Raccah et al., 2021; Seth & Bayne, 2022). However, what exactly what-it-is-likeness entails is a matter of extensive debate. For instance, there is extensive and ongoing discussion about whether conscious perception is rich or sparse (Block, 2011a, 2014; Knotts et al., 2019; Kouider et al., 2010), or whether there are levels or degrees of consciousness (Barra et al., 2020; Bayne et al., 2016; Overgaard & Overgaard, 2010). In fact, part of the allure of a possible future determination of which theory of consciousness accurately captures the phenomenon likely is that it helps us pin down the extension of the phenomenon i.e., informs us about its nature. In this light, determining which theory of consciousness is right[5] contributes to defining the phenomenon. Returning to the objection, what we are interested in here is theories about how and when things emerge, whether the thing is an apple, or an orange is of less importance. Put differently, we are interested in the emergence of "fruit", and even if one theory holds that the fruit that emerges is apples, while another posits oranges, both are theories about fruit. So — to re-iterate — we are interested in the first point in ontogenetic development where there is something it is like to be the individual and we readily concede that what extant theories take something-it-is-likeness to entail differs significantly (see e.g. Block, 2011b; Rosenthal, 2011; Weisberg, 2011 for a clearcut example of this). Irrespective of the extension and importance imbued on what-it-is-likeness by a given theory, as long as the theory endorses the ground assumption, the theory needs to account for O-emergence. In the present context, this clarification serves to highlight that it is possible to map and project a theory's stance on O-emergence independently of the specific conception of consciousness and what-it-is-likeness advanced by the theory's proponents.

The sixth clarification pertains to the role of O-emergence in the project of assessing and comparing theories of consciousness and the general work on the hard problem of consciousness (Chalmers, 1995). We see at least two ways in which mapping and projecting of where theories stand on O-emergence may contribute in this context. The first is that mapping what each theory of consciousness posits with respect to O-emergence affords us a set of timepoints which we can than pro-

[5] This phrasing is colloquial in the sense that it ignores the significant discussion in the philosophy of science about issues such as whether theories can be proven, the problem of induction, falsification and so forth. The point we want to make stands even if only something weaker than "proof" or "truth" is possible.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

ceed to investigate empirically[6]. The results of these investigations — in turn — will be part of the empirical evidence counting for or against a theory. More broadly, the mapping of theories gives us places (times) to look for O-emergence, which may be of benefit to the work on the hard problem by allowing us to approach it from a new angle i.e., by focusing on the (ontogenetic) origin of the phenomenon to be explained. The second way considering O-emergence may be of value to the project of assessing and comparing theories of consciousness is simply my illuminating which theories actually have proposed accounts of O-emergence, and which have not. On the assumption that one thinks an adequate theory of consciousness should explain O-emergence, whether a theory has an account of this — and the detail of any such account — is a relevant parameter when assessing and comparing theories.

3 Three questions of O-emergence

In addition to the clarifications offered above, we take for granted that the theories to be evaluated in light of O-emergence are at least[7] those who endorse the ground assumption, i.e. who believe that the brain is necessary for consciousness. With this in the background we next present three questions relevant to analyzing a theory's account of O-emergence.

3.1 Question 1: Conceptual clarity

The first question is how well defined the theory's position on O-emergence is on conceptual grounds. With "well defined" we here mean the amount of specificity and clarity there is to a theory's conceptual framework underpinning its explanation of O-emergence. For most theories, conceptual clarity on O-emergence is proportional to the conceptual clarity already present in the general conceptual framework of the theory. To illustrate, on the higher-order theories an individual

[6] Given the current lack of unequivocal biomarkers of consciousness (and especially ones that can be detected *in utero* provided the associated ethical and methodological constraints) this endeavor is currently not feasible. However, it would be imprudent to assume that such biomarkers will never be found, and recent progress has been made in this domain (Frohlich et al., 2023). In one sense, O-emergence may constitute the ideal scenario for a contrastive study of consciousness.

[7] Plausibly, any theory positing that consciousness exists has some variant of O-emergence question to answer. For instance, dualistic theories can provide adequate answers, even if they may amount to: 'consciousness never emerges, it was always there because the soul grounds consciousness, and the soul is eternal'. Even some theories that deny the existence of consciousness need to answer questions in the vein of O-emergence, either about linguistic practices, or about the origin and nature of any illusions we may have about consciousness (see e.g. Chalmers, 2018). Similarly, panpsychist theories in their answer to the so-called combination problem (Coleman, 2014; Goff, 2009) would need to tell us when in ontogenetic development micro-conscious or proto-conscious elements combine in the requisite way to form the distinctive human consciousness. See also footnote 2.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

is in a conscious state X if the individual has a higher-order representation of X (Rosenthal, 2012). From this, it can be taken to follow that O-emergence occurs when a higher-order state of the requisite kind occurs for the first time during ontogenetic development. Importantly, while conceptual clarity on O-emergence can often be derived in this way, it is not the case necessarily, and there may be significant variations depending on the particular wording of a theory. For instance, between Rosenthal's actualist HOT theory and the wide intrinsicality view of Rocco Gennaro (Gennaro, 2016).

3.2 Question 2: Empirical clarity

This question concerns the amount of clarity a theory delivers regarding O-emergence in the empirical domain, i.e. in terms of events in the brain (c.f. the ground assumption). We can subdivide empirical clarity into two aspects. First, topographical aspects, i.e. whether the explanation offered for O-emergence invokes specific brain areas. Secondly, whether it invokes various kinds of neural features, such as potentiation, plasticity, memory stores, neurotransmitters, connectivity, to name a few. To assess empirical clarity is to consider in conjunction the topographical claims and the feature claims of a theory. At issue is whether the conjunction sufficiently illuminates the posited transition from an unconscious individual to a conscious one. Now, clearly, clarity in the empirical domain is a tall order for all theories[8], and presently no theories give an (non-contentious) exhaustive answer. However, the fact that the question is really hard does not entail that we cannot evaluate and compare the quality of the answers offered by the various theories of consciousness.

3.3 Question 3: Physical, structural, and functional

The third question concerns the kind of changes in the brain that are *sufficient* for O-emergence. There seems to be three possible kinds of changes a theory can espouse: physical, structural, and functional. Holding that the operative factor underpinning O-emergence is physical consists in the view that certain brain areas X are sufficient for O-emergence, and once X are fully developed O-emergence occurs. Embracing the structural position consists in the view that specific structural properties are sufficient for O-emergence. One obvious candidate for such structural features could be the development of specific neural pathways, properties of such pathways (e.g. overall connectivity), or sets of pathways constituting connection hubs. Finally, the functional position consists in the view that the development of certain functional capacities is sufficient for O-emergence. An example of a candidate for a functional feature of this kind could be internal monitoring. Crudely put, one might say that the physical position concerns the development

[8] One obvious reason for this is that the question of O-emergence in relation to brain function has a close affinity to the notorious hard problem of consciousness.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

of brain areas, the structural position concerns relations between them, and the functional position concerns what brain areas (or sets of them) do.

3.4 Conclusions about O-emergence in development

It is not always obvious from a merely neural description of events what the implications of said events are at the level of personhood, and being clear about what a given theory entails with respect to when a human can be assumed to be conscious may have significant implications in a range of other domains (e.g. ethics; see Overgaard & Kirkeby-Hinrup, 2021b). For this reason, we will conclude the treatment of each theory by synthesizing and summing up the answers provided to the three questions. In doing this, we pay special attention to the consequences of its view on the (neurotypical) ontogenetic development of consciousness. Is the baby conscious prenatally? Is it conscious at the age 3 months? 6 months? 1 year? Or later? It is here important to reiterate that what the question concerns is *the earliest point at which phenomenal experiences occur*. As we explained in the introduction there *has* to be a before and an after.

4 SOMA and O-emergence

Since the proponents of SOMA highlight explicitly the importance of addressing O-emergence, SOMA is a suitable candidate to be the first theory to be evaluated using the three questions presented in the previous section. Because the proposal of SOMA is so recent (in terms of academic publishing time), there is — as of yet — not much literature on it. However, the original article along with the immediately following discussion nevertheless contain substantial claims and are sufficient to (preliminarily) assess SOMA in light of O-emergence. Certainly, given that the hypothesis is in its infancy, there is room for development and refinement which will allow its proponents to further flesh out the claims and implications of SOMA with respect to O-emergence. Ideally, any shortcomings or outstanding questions remaining after the treatment in this paper will serve to spur further refinement and development of the theory. With this caveat in mind, we now turn to assessing how SOMA handles O-emergence. Initially, it is worth noting that Cleeremans et al. explicitly (Cleeremans et al., 2020, p. 114) frame the development of SOMA in light of what we call O-emergence (as also revealed by the title "Learning to be conscious"). Their paper is explicit in the relevance and importance of what we call the ontogenetic emergence challenge, i.e. that a central (but often overlooked) question concerns the changes in the brain through development that initially brings about consciousness.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

4.1 Question 1: Conceptual clarity

The central claim of SOMA is that the emergence of consciousness depends on learning (Cleeremans et al., 2020. 115). The way SOMA deploys learning to explain the acquisition of consciousness depends on a set of interconnected claims. A core claim is that consciousness emerges as result of the development of a second-order system that observes and learns about the representations and dynamics of the network of first-order states the individual is in. In the words of Cleeremans *et al.* (Cleeremans et al., 2020, p. 117) this second-order system "subtends phenomenal experiences".

Another critical component of SOMA's explanation of O-emergence is the so-called self-other loop. The self-other loop consists in observations and interactions with others, and through these a model of agenthood is developed. This model is constructed from smaller models of unobservable internal states inferred in – or attributed to - other agents, by extrapolation from the knowledge the individual has of her own states. The main process driving the inference and attribution of unobservable inner states in other agents is the predictive processing involved in the perception-action loop (that precedes the self-other loop in ontogenetic development). Upon the development of a model of agenthood, this model is available for application to the individual herself. It is the application of models of agenthood to oneself that results in the emergence of a self and phenomenal experience (Cleeremans et al., 2020, p. 120). Thus, with respect to conceptual clarity, SOMA provides an internally consistent and very detailed account of the different stages in the process resulting in O-emergence.

4.2 Question 2: Empirical clarity

When it comes to empirical clarity, SOMA suggests a range of mechanisms and capacities that drive the acquisition of models of agenthood and handle the attribution of these to the individual herself, resulting in the individual learning to be conscious. The core mechanism underpinning the whole process is plasticity, i.e. the brain's ability to change and adapt dynamically over time. Thus, it is plasticity that bootstraps the implementation of the three loops central to SOMA. Briefly, in the first stage (the inner loop), the brain engages in self-monitoring to gather information about — and keep track of — which states it is in. Through informational integration processes, this results in a kind of (unconscious) metacognition, what Cleeremans and colleagues call a 'higher-order network' engaged in representational re-description of the first-order states (e.g. Cleeremans et al., 2020, p. 118 fig. 1). Through the deployment of predictive coding mechanisms, the individual next (in the perception-action loop) correlates execution of actions with perceptual input. Finally, (in the self-other loop) the individual deploys theory of mind capacities (that plausibly deploy knowledge acquired by the individual about herself via the inner-loop and the perception-action loop) to infer inner unobservable

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

states in other individuals, yielding the models of agenthood which can then be applied to herself, resulting in the acquisition of consciousness.

To summarize, the features invoked by SOMA in their explanation of O-emergence are plasticity, self-monitoring, metacognition, theory of mind, information integration, generative models, recurrent activity, and predictive coding. Each of these features are fairly well understood both theoretically and empirically, with the caveat that most are also the subject of significant debate.

If we turn our attention to the topographical aspect of empirical clarity, SOMA is less specific. In line with its reliance on ideas from global workspace, SOMA suggests that brain function cascades across the brain (Cleeremans et al., 2020, p. 114). In line with its affinity with the higher-order thought theories, proponents of SOMA consider the prefrontal cortex important, but remain somewhat noncommittal (Cleeremans et al., 2020, p. 117). They highlight the importance of the PFC in handling metacognition, monitoring and control, but nevertheless suggest that the models relevant to consciousness may be instantiated anywhere in the brain (Cleeremans et al., 2020, p. 118). One might try to determine the most likely topographical instantiation based on the close relationship to higher-order thought theories, global workspace (that both have advanced specific topographical areas or structures as relevant to their theory (e.g. Lau & Rosenthal, 2011; Mashour et al., 2020) posited by proponents of SOMA. One might also try to extrapolate topography based on the features suggested in SOMA to be involved in acquiring consciousness. For instance, both theory of mind (e.g. Frith & Frith, 2006; Gallagher & Frith, 2003) and metacognition (e.g. Fleming & Dolan, 2012) have generally been associated with specific topographical areas. Nevertheless, until further details are published by proponents of SOMA the topographical aspect of empirical clarity remains underspecified.

4.3 Question 3: Physical, structural, or functional

SOMA posits as the stimulus condition for O-emergence changes of the functional kind, i.e. something the brain becomes able to do. To illustrate this, one can point to the proponents of SOMA suggesting that (Cleeremans et al., 2020, p. 120): "*something unique* happens when a developing agent has models of itself available to it" (*italics* added) and later "*something special* happens when we try to build a model of the internal, unobservable states of agents that are just like ourselves" (Cleeremans et al., 2020, p. 121, *italics* added).[9]

[9] Some might think it is worth asking what exactly these quotes mean, and how they are supposed to explain the emergence of consciousness. Why should the brain learning to 1) re-describe its own representations to itself 2) predict the consequences of its actions on the world and 3) to re-describe its own activity to itself based on an (implicit, unconscious, enactive, embodied) internal model of agenthood derived from observation of other agents, yield a very special and unique type of knowledge or process that – when achieved – makes qualia appear and be continuously present in that individual from that moment on? However, questions of this kind apply to every theory, given that they tacitly invoke the famous explanatory gap (Levine, 1983). Thus, one might

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

4.4 Conclusions about O-emergence in development

SOMA is not clearly committed to a point in development for O-emergence. Cleeremans et al. (2020, p. 120) suggest that the acquisition of consciousness, and the development of a self, emerge progressively through an ongoing process. It is unclear if this suggestion amounts to the kind of graded process discussed above. In any case, it is possible to extrapolate possible implications regarding O-emergence from their current body of work. For instance, one might argue (as some have done; e.g. Overgaard & Kirkeby-Hinrup, 2021b) that the fact that the self-other loop relies on theory of mind processes seemingly entails that O-emergence cannot take place before the cognitive machinery catering to theory of mind has been developed. Furthermore, since there is evidence that fully fledged theory of mind does not develop until the age of 3 or 4 years, this suggests that an upshot of SOMA is that children below that age cannot be said to be phenomenally conscious. However, proponents of SOMA appear to reject this, albeit without giving an alternative answer (Cleeremans et al., 2021).

5 REFCON and O-emergence

REFCON is a theory of consciousness originating from the general neurocognitive theory with the moniker Reorganization of Elementary Functions (REF)[10]. REF was created to account for two seemingly contradictive phenomena. On the one hand, functional localization in neurotypical individuals and functional recovery after focal acquired brain injury on the other.

The REF model can roughly be described as a connectionist network within which the "units" are advanced processing modules called Elementary Functions (EFs)[11]. According to the REF model, all behavioral and/or mental phenomena are surface phenomena. This includes e.g. task solution, thoughts, and consciousness. The neural substrate of a given EF is localized within a restricted subdivision of a neural structure. Every EF performs fixed information processing on whatever input it is given and provides an output for further processing. Importantly, the information processing performed by individual EFs cannot be characterized according to traditional "functional/cognitive" terms. Rather, it can best be described

just as well ask how and why recurrent processing, higher-order representation, or integration of information could give rise to conscious experience. We certainly agree that such questions are important to ask. Nevertheless, given that the present objective is an explorative investigation of SOMA's account of O-emergence, these questions are not of relevance here, and will be left to others to address (Schurger & Graziano, 2022).

[10] Given its complexity, we cannot here provide a fully detailed introduction to the REF framework. Thus, we will limit ourselves to introducing the features relevant to assessing the questions of O-emergence and urge readers to consult the comprehensive literature available (Mogensen et al., 2018; Mogensen & Overgaard, 2018a, 2018b).

[11] This is in sharp contrast to traditional connectionist networks within which the "unit" is a "functionally indifferent" neuron.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

in mathematical terms. A given EF will typically simultaneously be involved in information processing associated with several traditional cognitive domains. Every structure/substructure of the brain (e.g. the hippocampus) mediates hundreds or thousands of unique EFs. According to the REF model, the "bridge" between the strictly localized and low-level information processing EFs and the surface phenomena (e.g. problem solving and/or mental phenomena) are the Algorithmic Strategies (ASs). ASs consist of numerous interacting EFs and are distributed in the sense that the neural substrate of an AS includes both the neural substrate of the individual EFs and the neural connections mediating the interactions between these EFs (Mogensen et al., 2018). Given the distributed nature of EFs, this means that typically, an AS will include EFs spanning numerous parts of the brain. Because an AS is the neurocognitive mechanism mediating a given surface phenomenon, this implies that the neural underpinnings of the surface phenomena are likewise distributed.

The original REF model primarily focused on the mechanisms of problem solving as well as basic neurocognitive processes. To account for perceptual processes (including conscious perception) REF was expanded with the REFCON model (Overgaard & Mogensen, 2014, 2015). The REFCON model is based on the same units and dynamic principles as the original REF model but introduces several new entities and concepts. The two most essential components are the Perceptual Algorithmic Modules (PAMs) and the Situational Algorithmic Strategy (SAS). Briefly: PAMs represent the external entities being perceived and are hierarchically ordered. The lower-level PAMs represent features rather than objects, while PAMs at progressively higher levels represent more complex entities and eventually individual objects. PAMs are selected in a "mutual competition" process – and the process of perceiving consists in the selection of PAMs of constantly higher levels. Fully identifying what is perceived is the selection of a PAM of the highest level. Importantly, PAMs (even those of the highest level) cannot in themselves mediate mental phenomena – including perceptual awareness. The way they contribute to surface phenomena is by integration into the SAS, making them available for consciousness and/or action.

The SAS is a highly specialized, dynamic network reflecting the current state of the individual. Thus, the SAS (which is distributed across practically all of the brain – including EFs from virtually all brain structures) represents not only the current perceptual situation but also the general internal – including mental – status of the individual. The use of refined methods such as the Perceptual Awareness Scale (PAS) (Ramsøy & Overgaard, 2004) has demonstrated that consciousness is best understood as a process that is present in degrees – spanning a number of intermediate steps from totally absent to totally present (Koch & Preuschoff, 2007; Overgaard et al., 2006). According to the REFCON model it is the degree of PAM integration into SAS that determines to what extent a perceptual entity is available for perceptual awareness. Thus, the perceptual process will – via selection of PAMs of progressively higher levels and eventual integration into SAS – result

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

in a SAS configuration with a more or less integrated PAM of the highest level representing what is being perceived.

5.1 Question 1: Conceptual clarity

Given that REFCON is originally conceived to answer a separate question, to get a handle on what it entails with respect to ontogenetic emergence, one needs to extrapolate from the fact that REFCON links consciousness to information that is available for action, in the form of algorithmic modules (AMs). While the exact type of action in question is still considered unknown (*personal communication*), the current hypothesis is that it is probably a special kind of flexible action. Furthermore, given that REFCON envisions mental phenomena such as thought and action as surface phenomena mediated by the SAS, REFCON indicates that O-emergence may be connected to the first appearance of these. We should not be surprised that REFCON does not deliver very detailed conceptual clarification of the processes involved in the ontogenetic acquisition of consciousness for three reasons. First, the theory was originally conceived to address a separate problem. Secondly, the central explanatory posits deployed in REFCON are mathematical, as opposed to conceptual. And finally, the questions of O-emergence as criteria for theories are only first presented in this paper and are therefore only now available as guidance for future development of the REFCON model. So, while we are unable to reach a conclusion on the conceptual clarity question for REFCON, we have nevertheless set a foundation and provided a frame for the debate on this.

5.2 Question 2: Empirical clarity

REFCON does not specify any single brain area related to consciousness since both the EFs, AMs and the SAS are distributed across the whole brain. Furthermore, because the REF framework allows for flexible reorganization, if REFCON eventually comes to associate a particular type of neural mechanism with integration into the SAS, this will likely merely reflect the neurotypical mechanism but *not the only way* integration into the SAS may be instantiated (*personal communication*). While REFCON — for the reasons just given — is underspecified in terms of localization, there are two other parameters on which REFCON shows empirical promise in relation to O-emergence. First of all, because the elements in REFCON are mathematical, the model has the potential to very precisely determine O-emergence at some point in the future, both in neurotypical development and on an individual level. However, the exact mathematics remain unspecified. Secondly, REFCON gives rise to a series of experimental predictions, which are useful to evaluate O-emergence. For instance, one prediction is that consciousness is graded rather than dichotomous, and that the gradedness correlates with performance.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

5.3 Question 3: Physical, structural, or functional

REFCON clearly advances a functional answer as consciousness is associated with information being integrated in SAS – independent of any physical structure. REFCON leaves room for the possibility that it may be the case that certain brain processes are necessary for consciousness, but this would be considered an empirical fact. Nothing in the REFCON theory requires that there is any physical structure that has a role of being necessary for consciousness.

5.4 Conclusions about O-emergence in development

REFCON does not specify one specific moment during the ontogenetic development where consciousness arises. Due to the resistance to associate any specific neural organization or structure with consciousness, an investigation into the physical development may be inconsequential. REFCON relates consciousness to information available for action – in particular flexible action. This aspect is underdefined in the theory, as it considers it a topic to be answered by future research exactly what may characterize such actions (*personal communication*). The idea is based on the simple observation that all things we are conscious of, we are able to speak about or act on – whether this be in a very simple sense or with full insight into the object or the action. Returning to O-emergence, it is possible to tentatively extrapolate an answer from REFCON. The extrapolation goes like this: the earliest movements are observed in the fetus around week 7, the neuromuscular junction is observed already in week 9, and at week 12, the peripheral nerves in the fetus have reached their muscle targets (Hayat & Rutherford, 2018). The neural infrastructure for sensory and proprioceptive feedback appears relatively early in gestation; classical experiments in postmortem fetal tissue demonstrated reflexive responses to tactile cutaneous stimulation very early in development (Hakamada et al., 1988). These findings complete the neural circuit from the cortex to the effector organs of movement at a relatively early stage of in utero development, which allows for spontaneous motor behavior to manifest. Although it is difficult for REFCON to commit to a specific moment in the ontogenesis for the appearance of consciousness, the theory tentatively proposes that a fetus between 7 and 12 weeks is conscious, if not even before (*personal communication*).

6 GNW and O-emergence

Unlike SOMA and REFCON, the Global Workspace theory, along with its younger sibling the Global Neuronal Workspace theory, has been around for many years. There has been ample work put into the development of the workspace model, and this is clearly reflected in how it handles the questions related to O-emergence. Yet, when it comes to O-emergence specifically, the workspace theories still provide only a relatively small body of work. Connections between Global Workspace the-

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

ory and O-emergence are mentioned in passing occasionally (see e.g. Changeux, 2017; Mashour et al., 2020), but only a few papers are devoted to the topic (Lagercrantz, 2009; Lagercrantz & Changeux, 2009).

The central idea in the global workspace theory is that consciousness depends critically on the availability of information. A mental state becomes conscious, when it is broadcast to a vast array of cognitive consumer systems. In the vernacular of the Global Neuronal Workspace this process is called ignition and is characterized by sudden non-linear activation profiles initiated in a (predominantly) frontoparietal network consisting of distributed and heavily interconnected neurons with long-range axons (Mashour et al., 2020). Activations in the global workspace are driven by so-called workspace neurons that selectively mobilize or suppress, through descending connections, the contribution of specific processor neurons (e.g. in sensory cortices) depending on task relevance.

6.1 Question 1: Conceptual clarity

In the two papers mentioned above (Lagercrantz, 2009; Lagercrantz & Changeux, 2009) specifically discussing GNW in relation to O-emergence, the authors provide relatively little conceptual clarification, instead relying on the large body of work already available on the workspace framework. Given the prominence of workspace models in debates about consciousness, and consequently the number of publications over the last couple of decades discussing this, we, like Lagercrantz and Changeux, will not rehash every aspect of the workspace models here, but merely note that on the question of conceptual clarity, the GNW is on very solid ground given that the answer piggybacks on the existing conceptual framework. In brief, consciousness depends on broadcasting (ignition), and in relation to O-emergence what matters is when the sufficient conditions for this come about.

6.2 Question 2: Empirical clarity

Things get more interesting when we turn to the second question about the empirical clarity of GNW in relation to O-emergence. This is where the importance of the contributions of Lagercrantz and Changeux becomes apparent. Lagercrantz (2009) posits that (neurotypical) newborns are clearly conscious according to all ten criteria Lagercrantz deploys to assess the presence of consciousness (importantly, the criteria mentioned here should not be conflated with the ones proposed by Doerig et al. discussed above. For details on Lagercrantz's criteria see his 2009, pp. 57–58). However, the question of whether babies may be conscious prenatally remains. Lagercrantz and Changeux explore this question indirectly through discussion of the possibility of consciousness in very preterm babies *ex utero*. Initially, in line with the ground assumption discussed in the introduction, a clear cut-off before which the authors rule out the existence of consciousness can be established based on the initial development of the requisite neural structures underpinning the global workspace. Lagercrantz (Lagercrantz, 2009, p. 58) sets this cutoff

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

around the 23-24th week of pregnancy referring to the fact that it is only around this time that the thalamocortical connections are established. In the Changeux paper (Lagercrantz & Changeux, 2009, p. 259), the authors posit that before the 26th week of pregnancy, the immaturity of the brain networks is such that it is implausible that they can sustain 'minimal consciousness'. Since their treatment is based on preterm babies ex utero, it is unclear whether the authors hold that — prenatally — babies are entirely unconscious, and that consciousness appears at birth. One reason this is unclear is that they state that "in utero the fetus is mostly in a state of unconsciousness" (Lagercrantz & Changeux, 2009, p. 258). In relation to O-emergence, the presence of the word "mostly" in this quote is of central concern. This is because 'mostly' indicates that the fetus *in utero* may be intermittently conscious i.e., a case of the kind of intermittent bursts of consciousness discussed in clarification two of the introduction above. If the fetus *in utero* is intermittently conscious then GNW cannot hold that O-emergence coincides with birth. On the other hand, they seem to suggest a role for birth in O-emergence through suggesting that behavioral signs in utero seem to be preprogrammed and of subcortical origin, that the fetus is largely sedated and that noxious stimuli cause inhibition instead of arousal (ibid). Continuing to suggest that arousal incurred by multiple distinct causes at birth may boot up, as it were, consciousness. This is supposedly achieved at birth by stress activation of the cholinergic system, the sudden arrival of a host of new sensory stimuli, and the removal of endogenous constraints on consciousness (e.g. sedation in the uterus by pregnanolone and prostaglandin). In any case, we will remain neutral on this ambiguity here and merely note that there seems to be at least two options for a proponent of GNW. The first option is to affix O-emergence to birth. The second option is to hold that in utero the fetus is conscious (for at least a single point in time prior to birth). We look forward to a clarification of this from proponents of GNW.

6.3 Question 3: Physical, structural, or functional

The answer to this question depends on which stance is taken on the role of birth just discussed. If one takes the position that birth is the key factor, then the changes relevant to O-emergence are of the physical and are the ones associated with birth. If one does not think birth is the key factor, then it is unclear what the relevant changes are. If one thinks the full consolidation of a global workspace at some point *in utero* is sufficient, then that would be a reason to give a structural answer. Importantly, both options share a necessary condition for O-emergence which is the development of the requisite neural architecture to support the neuronal workspace. This necessary condition is in place around the 24-26th weeks of pregnancy, which means that regardless of which position one takes on birth, O-emergence cannot occur before that point.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

6.4 Conclusions about O-emergence in development

The answer from Lagercrantz and Changeux seems to be that consciousness emerges at birth, with the caveat that (as we saw in their answer to the second question) that preterm babies born before the 24th to 26th weeks are likely to lack the requisite neural development to sustain consciousness. If one thinks that birth is the key factor, one interesting implication is that after the 26th week of pregnancy the fetus seemingly is counterfactually conscious, in the sense that (ceteris paribus) *if* it were to be born it *would* become conscious. Now, one might object that for the GNW account what really matters is broadcasting (ignition) and that birth is irrelevant. We agree that on the theoretical level, this seems to be the right answer. However, in the discussion of O-emergence, what we are interested in is the empirical reality, i.e. when consciousness typically emergences in actual (neurotypical) human development. This perspective is also the one taken by Lagercrantz and Changeux, whose work we have relied on here. So while we acknowledge that there is a theoretical answer, which is that O-emergence coincides with the first instance of broadcasting, what we are interested in here is the empirical aspect which is *when* does the first instance of broadcasting typically occur. Disregarding the outstanding questions we have highlighted in relation to the views of Lagercrantz and Changeux, their answer appears to be that it occurs at birth.

7 Concluding remarks

In the introduction, we applauded the extant efforts to assess and compare theories of consciousness. The aim of this paper has been to complement these efforts by introducing a new parameter on which to assess theories of consciousness. We took as a starting point the assumption that consciousness depends on brain activity, and from this concluded that since the brain does not yet exist at conception, there has to be a point in ontogenetic development at which consciousness emerges in an individual. Given that the transition from an *in esse* unconscious being to a conscious one can be considered an important turning point in the ontogenetic development of an individual, we were surprised by the scarcity of treatment ontogenetic emergence has been given in the debates between competing theories of consciousness. To remedy this, we proposed three questions to assess where a given theory stands on O-emergence. Then, we applied these criteria to three theories of consciousness to assess the extent to which each could account for O-emergence. This served to provide a foundation and frame for the theoretical side of a debate on O-emergence. It is worth noting however that on the empirical side of these questions, recent work (Frohlich et al., 2023) on biophysical markers of consciousness measurable in utero looks promising.

Finally, a few things are worth underscoring in relation to the investigation of O-emergence presented here. The first thing is merely a reiteration of the fact

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

that O-emergence is — as of yet — an underexplored domain within the debates on theories of consciousness. An upshot of this is that even theories which actually have considered O-emergence (e.g. SOMA and GNW) still have significant outstanding questions, and naturally, theories that have not yet treated questions of O-emergence (e.g. REFCON) have even more questions to answer in this regard. Therefore, a significant body of work remains across the field of candidate theories to develop accounts of O-emergence.

The second thing concerns the context of this investigation. The context is the overarching project of assessing and comparing theories of consciousness. In this project O-emergence is but one of several criteria. However, outside of this context, addressing O-emergence belongs on the to-do list of proponents of any theory of consciousness that endorse the ground assumption (that consciousness depends on brain activity). At some point in the transition from an entity with *no brain and therefore no consciousness* to a human individual that has a brain and is conscious, consciousness must appear for the first time. Given that O-emergence follows from the ground assumption, and moreover concerns the very genesis of phenomenon that is the central explanandum of theories of consciousness, it may be considered a feature of the phenomenon that a successful theory ought to account for.

The third thing is the approach we have taken here. Even if there is a large body of work yet to be done on O-emergence, the tools to take it on are available. We have offered three questions that can be deployed to evaluate a theory in light of O-emergence. We have suggested that answers to one or more of these questions sometimes can be derived from the general theoretical and empirical frameworks of a theory. This has two important implications. First, theories that have not yet explicitly treated O-emergence do not need to start from scratch. Secondly, it is possible to assess where a theory stands on O-emergence even if its proponents have not addressed this explicitly. Explicit answers from a theory's proponents (of course) are preferable to extrapolating answers to questions of O-emergence from the theory's general theoretical and empirical frameworks. However, the latter may serve to scaffold the former by drawing attention to apparent implications or problems that need to be addressed by a theory in relation to O-emergence. This may serve to guide and inspire work on O-emergence, thus opening future avenues of research. Examples of this are available from all three cases we have treated here. To elaborate, in our analysis of SOMA we highlighted an implication of SOMAs dependence on theory of mind (ToM). Because ToM is only fully developed around three to four years of age, this seems to entail that O-emergence occurs correspondingly late. If something less than fully fledged ToM is sufficient for O-emergence, interesting questions arise regarding what it is, why it is so, and how it fits into the rest of the SOMA framework. In our analysis of GNW we identified a need for elaboration on two aspects. The first concerned the phrase "in utero the fetus is mostly in a state of unconsciousness" (Lagercrantz & Changeux, 2009, p. 258). Does this mean that the fetus is *sometimes* conscious in utero? The answer to this question is important to understanding GNW's account of O-emergence.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

The second aspect in need of elaboration by proponents of GNW concerns how to understand the seemingly counterfactual ascription of consciousness that follows from approaching the issue by asking *when* a fetus *would* be conscious *if* it was born. Our analysis of REFCON illustrates how much can be derived from the general theoretical and empirical framework of a theory, and how this may guide future explicit treatment of O-emergence as well as the overall development of a theory.

References

Albantakis, L. (2020). Integrated information theory. In *Beyond neural correlates of consciousness* (pp. 87–103). Routledge.

Barra, A., Carrière, M., Laureys, S., & Martial, C. (2020). From unconscious to conscious: A spectrum of states. In *Beyond neural correlates of consciousness* (pp. 16–43). Routledge.

Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are there levels of consciousness? *Trends in Cognitive Sciences*, *20*(6), 405–413.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5-6), 481-99; discussion 499-548. https://doi.org/10.1017/S0140525X07002786

Block, N. (2011a). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567–575. https://doi.org/10.1016/j.tics.2011.11.001

Block, N. (2011b). The higher order approach to consciousness is defunct. *Analysis*, *71*(3), 419–431. https://doi.org/10.1093/analys/anr037

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences*, *18*(9), 445–447.

Blum, L., & Blum, M. (2022). A theory of consciousness from a theoretical computer science perspective: Insights from the conscious turing machine. *Proceedings of the National Academy of Sciences*, *119*(21), e2115934119. https://doi.org/doi:10.1073/pnas.2115934119

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. https://doi.org/10.1016/j.tics.2019.06.009

Cabanac, M., Cabanac, A. J., & Parent, A. (2009). The emergence of consciousness in phylogeny. *Behavioural Brain Research*, *198*(2), 267–272.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, *25*(9-10), 6–61.

Changeux, J.-P. (2017). Climbing brain levels of organisation from genes to consciousness. *Trends in Cognitive Sciences*, *21*(3), 168–181.

Ciaunica, A., Constant, A., Preissl, H., & Fotopoulou, K. (2021). The first prior: From co-embodiment to co-homeostasis in early life. *Consciousness and Cognition*, *91*, 103117. https://doi.org/https://doi.org/10.1016/j.concog.2021.103117

Ciaunica, A., Safron, A., & Delafield-Butt, J. (2021). Back to square one: The bodily roots of conscious experiences in early life. *Neuroscience of Consciousness*, *2021*(2). https://doi.org/10.1093/nc/niab037

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., Vuillaume, L., & De Heering, A. (2020). Learning to be conscious. *Trends in Cognitive Sciences*, *24*(2), 112–123.

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., Vuillaume, L., & De Heering, A. (2021). Do you need to be conscious to learn to be conscious? *Trends in Cognitive Sciences*, *25*(1), 9–11.

Coleman, S. (2014). The real combination problem: Panpsychism, micro-subjects, and emergence. *Erkenntnis*, *79*(1), 19–44. https://doi.org/10.1007/s10670-013-9431-x

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. https://doi.org/http://dx.doi.org/10.1016/S0010-0277(00)00123-2

Delafield-Butt, J. T., & Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review*, *33*(4), 399–425. https://doi.org/https://doi.org/10.1016/j.dr.2013.09.001

Delafield-Butt, J., & Trevarthen, C. (2022). Consciousness generates agent action. *Behavioral and Brain Sciences*, *45*, e44. https://doi.org/10.1017/S0140525X2100203X

Doerig, A., Schurger, A., & Herzog, M. H. (2020). Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience*, *12*(2), 41–62.

Fink, S. B. (2016). A deeper look at the "neural correlate of consciousness." *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01044

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frohlich, J., Bayne, T., DallaVecchia, A., Kirkeby-Hinrup, A., Mediano, P. A., Moser, J., Talar, K., Gharabaghi, A., & Preissl, H. (2023). Not with a "zap" but with a "beep": Measuring the origins of perinatal experience. *Neuroimage*, *120057*.

Frohlich, J., Toker, D., & Monti, M. M. (2021). Consciousness among delta waves: A paradox? *Brain*, *144*(8), 2257–2277.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences*, *7*(2), 77–83.

Gennaro, R. J. (2016). Higher-order thoughts, neural realization, and the metaphysics of consciousness. In *Consciousness: Integrating eastern and western perspectives* (pp. 83–102). New Age Publishers.

Goff, P. (2009). Why panpsychism doesn't help us explain consciousness. *Dialectica*, *63*(3), 289–311.

Hakamada, S., Hayakawa, F., Kuno, K., & Tanaka, R. (1988). Development of the monosynaptic reflex pathway in the human spinal cord. *Developmental Brain Research*, *42*(2), 239–246.

Hayat, T. T., & Rutherford, M. A. (2018). Neuroimaging perspectives on fetal motor behavior. *Neuroscience & Biobehavioral Reviews*, *92*, 390–401.

Kirkeby-Hinrup, A. (2024). Interdisciplinary consciousness studies needs philosophers of science. *Filosofiska Notiser*, *11*(1), 3–18.

Kirkeby-Hinrup, A., & Fazekas, P. (2021). Consciousness and inference to the best explanation: Compiling empirical evidence supporting the access-phenomenal distinction and the overflow hypothesis. *Consciousness and Cognition*, *94*, 103173.

Knotts, J., Odegaard, B., Lau, H., & Rosenthal, D. (2019). Subjective inflation: Phenomenology's get-rich-quick scheme. *Current Opinion in Psychology*, *29*, 49–55.

Koch, C., & Preuschoff, K. (2007). Betting the house on consciousness. *Nature Neuroscience*, *10*(2), 140–141.

Kouider, S., De Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, *14*(7), 301–307.

Lagercrantz, H. (2009). The birth of consciousness. *Early Human Development*, *85*(10), S57–S58.

Lagercrantz, H., & Changeux, J.-P. (2009). The emergence of human consciousness: From fetal to neonatal life. *Pediatric Research*, *65*(3), 255–260.

Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks*, *17*(5-6), 861–872. https://doi.org/10.1016/j.neunet.2004.02.005

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Levine, J. (1983). Materialism and qualia, the explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354–361. <Go to ISI>://WOS:A1983SU49200007

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798.

Mogensen, J., Daugaard, N., Kitsios, S., Pedersen, J. E., & Overgaard, M. (2018). Understanding the neurocognitive organization as strategies rather than functions: Implications for neurological research. *Neurology*, *10*(11), 1008–1016.

Mogensen, J., & Overgaard, M. (2018a). Neural connections and mental states: The need for a neurocognitive framework. *Ec Neurology*, *10*, 180–194.

Mogensen, J., & Overgaard, M. (2018b). Reorganization of the connectivity between elementary functions as a common mechanism of phenomenal consciousness and working memory: From functions to strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170346. https://doi.org/10.1098/rstb.2017.0346

Mudrik, L., Mylopoulos, M., Negro, N., & Schurger, A. (2023). Theories of consciousness and a life worth living. *Current Opinion in Behavioral Sciences*, *53*, 101299. https://doi.org/https://doi.org/10.1016/j.cobeha.2023.101299

Nagel, T. (1974). What is it like to be a bat. *Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914

Northoff, G., & Lamme, V. (2020). Neural signs and mechanisms of consciousness: Is there a potential convergence of theories of consciousness in sight? *Neuroscience & Biobehavioral Reviews*, *118*, 568–587.

Overgaard, M., & Kirkeby-Hinrup, A. (2021a). Finding the neural correlates of consciousness will not solve all our problems. *Philosophy and the Mind Sciences*, *2*.

Overgaard, M., & Kirkeby-Hinrup, A. (2021b). Is learning the cognitive basis of consciousness? The moral implications of SOMA. *Trends in Cognitive Sciences*, *25*(1), 8–9.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902

Overgaard, M., & Mogensen, J. (2014). Visual perception from the perspective of a representational, non-reductionistic, level-dependent account of perception and conscious awareness. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1641), 20130209.

Overgaard, M., & Mogensen, J. (2015). Reconciling current approaches to blindsight. *Consciousness and Cognition*, *32*, 33–40.

Overgaard, M., & Overgaard, R. (2010). Neural correlates of contents and levels of consciousness. *Frontiers in Psychology*, *1*, 164.

Overgaard, M., Rote, J., Mouridsen, K., & Ramsøy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Consciousness and Cognition*, *15*(4), 700–708.

Passos-Ferreira, C. (forthcoming). Are infants conscious? *Philosophical Perspectives*. Forthcoming.

Raccah, O., Block, N., & Fox, K. C. (2021). Does the prefrontal cortex play an essential role in consciousness? Insights from intracranial electrical stimulation of the human brain. *Journal of Neuroscience*, *41*(10), 2076–2087.

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*(1), 1–23.

Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia*, *46*(3), 829–840. https://doi.org/10.1016/j.neuropsychologia.2007.11.012

Rosenthal, D. M. (2011). Exaggerated reports: Reply to block. *Analysis*, *71*(3), 431–437.

Rosenthal, D. M. (2012). Higher-order awareness, misrepresentation and function. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1594), 1424–1438. https://doi.org/10.1098/rstb.2011.0353

Schurger, A., & Graziano, M. (2022). Consciousness explained or described? *Neuroscience of Consciousness*, *2022*(1). https://doi.org/10.1093/nc/niac001

Seth, A., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*(7), 439–452.

Tononi, G. (2005). Consciousness, information integration, and the brain. In L. Steven (Ed.), *Progress in brain research: Vols. Volume 150* (pp. 109–126). Elsevier. https://doi.org/http://dx.doi.org/10.1016/S0079-6123(05)50009-8

Weisberg, J. (2011). Abusing the notion of what-it's-like-ness: A response to block. *Analysis*, *71*(3), 438–443.

Kirkeby-Hinrup, A., & Overgaard, M. (2023). Ontogenetic emergence as a criterion for theories of consciousness: Comparing GNW, SOMA, and REFCON. *Philosophy and the Mind Sciences*, *4*, 29. https://doi.org/10.33735/phimisci.2023.9902